

Главный редактор
д.т.н., проф., академик РАЕН
Щербаков А.Ю.

Ученый секретарь Редакционного совета
Рязанова А.А.

Верстка Мотова Н.В.

Издание включено в перечень ВАК (специальности: 2.3.2, 2.3.6, 2.3.8, 5.2.4)

ВЕСТНИК СОВРЕМЕННЫХ ЦИФРОВЫХ ТЕХНОЛОГИЙ

НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ



№25
ДЕКАБРЬ 2025
ISSN 2686-9373
www.c3da.org



Издатели: Российский государственный социальный университет
Ассоциация РКЦФА

Адрес редакции и издателя: 129226, Москва,
ул. Вильгельма Пика, д.4, стр.1
www.c3da.org

Подписано в печать 25.12.2025 г.
Тираж 500 экз.
Подписной индекс в каталоге «Пресса России»: 79111



Свидетельство о регистрации СМИ
ПИ № ФС 77-76187 от 08.07.2019 г.

Журнал включен в перечень рецензируемых научных изданий ВАК, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук.
(2.3.2) Вычислительные системы и их элементы
(2.3.6) Методы и системы защиты информации, информационная безопасность
(2.3.8) Информатика и информационные процессы
(5.2.4) Финансы

РЕДАКЦИОННЫЙ СОВЕТ

Главный редактор – Щербаков Андрей Юрьевич, доктор технических наук, профессор, заведующий кафедрой когнитивно-аналитических и нейро-прикладных технологий РГСУ, президент Ассоциации специалистов в области развития криптовалют и цифровых финансовых активов (Ассоциации РКЦФА).

Председатель Редакционного Совета – Сигов Александр Сергеевич, академик Российской академии наук, доктор физико-математических наук, член Научного совета при Совете Безопасности РФ, президент Российского технологического университета МИРЭА, заслуженный деятель науки Российской Федерации, почётный работник высшего профессионального образования РФ.

Сопредседатель Редакционного Совета – Хазин Андрей Леонидович, ректор Российского государственного социального университета, академик Российской академии художеств.

Сопредседатель Редакционного Совета – Елизаров Георгий Сергеевич, доктор технических наук, директор ФГУП «НИИ «Квант», академик Академии Криптографии РФ.

Ученый секретарь Редакционного Совета – Рязанова Алина Александровна, вице-президент Ассоциации РКЦФА по международному сотрудничеству, ведущий специалист Научно-образовательного центра социальной аналитики Российского государственного социального университета.

Запечников Сергей Владимирович, доктор технических наук, доцент, профессор Института интеллектуальных кибернетических систем Национального исследовательского ядерного университета «МИФИ», Вице-президент Ассоциации РКЦФА по научной работе.

Кириченко Татьяна Витальевна, доктор экономических наук, профессор, заместитель заведующего кафедрой безопасности цифровой экономики РГУ нефти и газа (НИУ) имени И.М. Губкина.

Князев Александр Викторович, доктор физико-математических наук, профессор, директор Института точной механики и вычислительной техники им. С.А.Лебедева.

Комзолов Алексей Алексеевич, доктор экономических наук, профессор, заведующий кафедрой безопасности цифровой экономики РГУ нефти и газа (НИУ) имени И.М. Губкина.

Конявский Валерий Аркадьевич, доктор технических наук, заведующий кафедрой Московского физико-технического института (МФТИ).

Новиков Владимир Геннадьевич, доктор экономических наук, доктор социологических наук, профессор, член-корреспондент РАН, советник ректора Российского государственного социального университета.

Сенаторов Михаил Юрьевич, доктор технических наук, профессор, действительный член Российской Академии космонавтики им. К.Э.Циолковского, почетный эксперт Ассоциации РКЦФА, президент Ассоциации инженерных компаний «Ситэс-Центр».

Шилова Евгения Витальевна, доктор экономических наук, профессор кафедры экономики знания Высшей школы современных социальных наук МГУ имени М.В. Ломоносова.

Алиев Джомарт Фазылович, доктор философии в области бизнес-права (PhD), доктор делового администрирования в области финансов (DBA), кандидат экономических наук, член-корреспондент Российской академии художеств.

Егоров Владимир Ильич, кандидат физико-математических наук, заместитель директора Национального центра квантового интернета.

Мачихин Дмитрий Сергеевич, эксперт по вопросам противодействия отмыванию доходов и финансированию терроризма (ПОД/ФТ), учета и комплаенса цифровых финансовых активов и валют, член профильного комитета при Государственной Думе РФ.

Правиков Дмитрий Игоревич, кандидат технических наук, заведующий кафедрой комплексной безопасности критически важных объектов РГУ нефти и газа (НИУ) имени И.М. Губкина.

Терпугов Артем Евгеньевич, кандидат экономических наук, Проректор Государственного университета управления.

РЕДАКЦИОННОЕ ПРИМЕЧАНИЕ

Завершающий 2025-й год, двадцать пятый выпуск нашего журнала посвящен важным для экономического и социального развития современного общества аспектам, прежде всего – использованию технологий искусственного интеллекта в таких областях, как медицина, добыча промышленность и энергетика.

Выпуск открывает статья **«Проблемы внедрения больших языковых моделей в медицине»** коллектива авторов, подготовленная с участием авторитетных экспертов в области медицины и искусственного интеллекта, в которой затрагивается весьма резонансный и актуальный для всей области современного здравоохранения вопрос готовности к внедрению языковых моделей в практической медицине. Рассмотренная зарубежная практика показала, что классическое тестирование с использованием стандартизованных тестов не является достаточно корректным, но применение специализированного набора стресс-тестов позволяет выявлять ошибки в диагностике. Исследование имеет важное значение для дальнейшего целеполагания в области медицинского искусственного интеллекта.

Внедрению технологий искусственного интеллекта в угледобывающую отрасль посвящена статья **«Создание цифрового двойника угледобывающего предприятия на основе геопространственных данных»**, также представленная большим коллективом авторов. Высокое качество и точность геопространственных данных, получаемых методами, описанными в статье, позволяет создавать актуальную горно-графическую документацию, строить цифровые модели и двойники. Использование цифровых моделей способствует снижению затрат, минимизации человеческого фактора, повышению безопасности, эффективности и производительности шахтных работ. Собственные эмпирические данные и оригинальные результаты интеграции цифрового двойника в web-приложение делают данное исследование прикладным и уникальным.

В статье **«Особенности удаленного видеомониторинга энергетической инфраструктуры арктической зоны»**, представленной коллегами из Государственного университета управления, проводится аналитический обзор энергетической инфраструктуры арктической зоны, рассмотрены виды объектов, подверженные наибольшему износу и внешним воздействиям, и систематизированы современные подходы к их мониторингу. Проанализированы архитектуры сверточных нейронных сетей, применяемых для детектирования, сегментации и классификации дефектов, определены специфические требования к построению обучающих выборок. Рассмотренные принципы, методы и технические решения создают основу для формирования комплексного мониторинга объектов в Арктике, обеспечивая повышение достоверности оценки состояния инфраструктуры, снижение вероятности аварийных ситуаций и оптимизацию эксплуатационных процессов.

Статья **«Аппаратный ускоритель вывода глубоких нейронных сетей на основе СФ-блоков блочной ОЗУ со встроенными вычислительными ядрами»** тандема авторов посвящена разработке архитектуры аппаратного ускорителя вывода глубоких нейронных сетей на программируемых логических интегральных схемах, основанной на сложных функциональных блоках блочной оперативной памяти со встроенными вычислительными ядрами. Предложенный подход реализует парадигму вычислений в памяти и направлен на снижение объёма обмена данными между памятью и вычислительной логикой. Описана гибридная организация вычислительной системы, в которой предлагаемые блоки выполняют массовые операции умножения-накопления, а блоки цифровой обработки сигналов обеспечивают операции повышенной точности и объединение результатов. Результаты моделирования подтверждают повышение пиковой производительности по сравнению с существующими архитектурами вычислений.

В работе **«Программные решения для анализа данных в правоохранительной деятельности»** тандема авторов исследуется состояние информационно-аналитического обеспечения правоохранительных органов в условиях цифровизации общества. На основании анализа накопленного опыта использования автоматизированных информационных систем сделан вывод об актуальной необходимости внедрения интегрированных программных платформ. Сформулирован необходимый набор функций программной платформы и обоснован выбор конкретного программного решения.

В новой статье Игоря Кириллова **«Математические аспекты симметрии в изобразительном искусстве»** обсуждаются математические аспекты понятия симметрии применительно к орнаментальному искусству, расцвет которого пришелся на XIII-XV вв. (мавританская Испания). Высказана интересная точка зрения на влияние мавританских орнаментов на современное состояние математической теории групп.

В выпуске также опубликован интересный материал Михаила Масленникова **«Россия и RSA. Почувствуйте разницу»**, касающийся истории и перспектив развития гражданской криптографии в нашей стране.

Редакционная **рецензия на книгу «Начала технической защиты информации»** Светланы Конявской-Счастной может помочь нашим читателям в выборе учебных материалов для работы в области обучения и дидактики.

Новогодний номер журнала завершает рассказ нашего постоянного автора Егора Федорова **«Неизбежный маршрут»**. Читатель, вероятно, уже привык к неожиданному и подчас более глубокому взгляду автора на явления привычные и кажущиеся общеизвестными, но этот рассказ способен не только удивить, но и побудить серьезно задуматься о морально-этических истоках изменений в масштабе целых социально-политических формаций.

СОДЕРЖАНИЕ

1. ФУНДАМЕНТАЛЬНЫЕ ПРОБЛЕМЫ ЦИФРОВЫХ ТЕХНОЛОГИЙ

- О.Н. Андреева, А.В. Домашев, Е.А. Евланова, А.А. Рязанова, А.Ю. Щербаков** – Проблемы внедрения больших языковых моделей в медицине 5
O.N. Andreeva, A.V. Domashev, E.A. Evlanova, A.A. Ryazanova, A.Yu. Shcherbakov – Problems of implementing large language models in medicine 5

2. ПРАКТИЧЕСКИЕ АСПЕКТЫ ЦИФРОВЫХ ТЕХНОЛОГИЙ

- Т.В. Чаплыгина, К.И. Петерс, Р.Е. Зверев, Е.С. Громов, С.А. Иванова** – Создание цифрового двойника угледобывающего предприятия на основе геопространственных данных 19
T.V. Chaplygina, K.I. Peters, R.E. Zverev, E.S. Gromov, S.A. Ivanova – Creation of a digital model twin of a coal mining enterprise based on geospatial data 19
- Д.В. Сердечный, Н.С. Акиньшин, В.Д. Кутков** – Особенности удаленного видеомониторинга энергетической инфраструктуры арктической зоны 29
D.V. Serdechnyj, N.S. Akinshin, V.D. Kutkov – Features of remote video monitoring of energy infrastructure in the arctic zone 29
- О.В. Зобов, А.А. Спиридонов** – Аппаратный ускоритель вывода глубоких нейронных сетей на основе СФ-блоков блочной ОЗУ со встроенными вычислительными ядрами 40
O.V. Zobov, A.A. Spiridonov – Hardware accelerator for deep neural network inference based on SF blocks of block RAM with integrated computing cores 40

3. СОВРЕМЕННЫЕ ЦИФРОВЫЕ ТЕХНОЛОГИИ: ОБЗОРЫ, МНЕНИЯ, ДИСКУССИИ

- В.Л. Акапьев, А.В. Борисенко** – Программные решения для анализа данных в правоохранительной деятельности 55
V.L. Akap'yev, A.V. Borisenko – Software solutions for data analysis in law enforcement 55
- Е.С. Гончаренко, И.А. Кириллов** – Математические аспекты симметрии в изобразительном искусстве 63
E.S. Goncharenko, I.A. Kirillov – Mathematical aspects of symmetry in the visual arts 63
- Михаил Масленников** – «Россия и RSA. Почувствуйте разницу» 72
- От редакции «Начать сначала»** 87

5. ЛИТЕРАТУРА О ЦИФРОВЫХ ТЕХНОЛОГИЯХ

- Егор Федоров** – Неизбежный маршрут 88

УДК: 004.8

Проблемы внедрения больших языковых моделей в медицине

O.N. Andreeva, A.V. Domashev, E.A. Evlanova,
A.A. Ryazanova, A.Yu. Shcherbakov

Problems of Implementing Large Language Models in Medicine

Abstract. This article analyzes the appliance readiness of medical artificial intelligence systems based on large multimodal language models, using international publications. Classic testing using standardized tests proves insufficiently accurate, while the use of a set of extended tests (stress tests) allows for the identification of diagnostic errors. It is shown that different test datasets actually test different qualities, and a methodology for their thematic "stratification" is presented. It is noted that average system ratings can create the illusion of progress and mislead about the models' readiness for real-world use. An idea is proposed for incorporating the patient's medical record or medical history to combat hallucinations in diagnosis. The main conclusion of the article is that the results demonstrated on medical benchmarks cannot be considered as a direct indicator of AI readiness for clinical implementation.

Keywords: diagnostics, language model, benchmark, stress test, medical artificial intelligence.

Российского государственного социального университета

О.Н. Андреева¹
А.В. Домашев²
Е.А. Евланова³
А.А. Рязанова⁴
А.Ю. Щербаков⁵

¹Кандидат медицинских наук, врач телемедицинских консультаций, заведующая гинекологическим кабинетом ООО «Доктор 2000» (партнёр сети медицинских центров «MedSwiss»).

E-mail: and-oksy@yandex.ru

²Руководитель Инженерного центра «ИНФО-СТАР» Академического института виртуальной и дополненной реальности Российской государственного социального университета.

³Аналитик Научно-образовательного центра социальной аналитики Российского государственного социального университета.

⁴Ведущий специалист Научно-образовательного центра социальной аналитики Российского государственного социального университета.

E-mail: a.ryazanova@c3da.org

⁵Доктор технических наук, профессор, заведующий кафедрой когнитивно-аналитических и нейро-прикладных технологий Российской государственной социальной университета, ведущий научный сотрудник Государственного университета управления.

E-mail: x509@ras.ru

Аннотация. В статье на материале зарубежных публикаций анализируется готовность к эксплуатации медицинских систем искусственного интеллекта, построенных на основе больших мультимодальных языковых моделей. Классическое тестирование с использованием стандартизованных тестов оказывается недостаточно корректным, а применение набора расширенных тестов (стресс-тестов) позволяет выявлять ошибки в диагностике. Показано, что разные тестовые наборы данных на самом деле проверяют разные качества, и изложена методология для их тематического «раслоения». Отмечается, что усреднённые рейтинги систем могут создавать иллюзию прогресса и вводить в заблуждение относительно готовности моделей к реальной эксплуатации. Предложена идея учета медицинской карты или медицинской истории больного для борьбы с галлюцинированием в диагнозе. Основной вывод статьи состоит в том, что результаты, продемонстрированные на медицинских бенчмарках, нельзя рассматривать как прямой показатель готовности ИИ к клиническому внедрению.

Ключевые слова: диагностика, языковая модель, бенчмарк, стресс-тест, медицинский искусственный интеллект.

ВВЕДЕНИЕ

В Национальной стратегии развития искусственного интеллекта на период до 2030 года в Российской Федерации указано, что цифровизация здравоохранения – одна из важнейших задач для получения качественных и быстрых медицинских услуг населением [1].

Однако внедрение систем искусственного интеллекта (ИИ) в медицине проходит не слишком быстрыми темпами (например, в Москве в настоящее только около 50 компаний применяют ИИ в

своей практике), поскольку велики риски в случае ошибочного использования, в первую очередь при постановке диагнозов.

В мире в целом отмечается достаточно активный рост инвестиционных вложений в развитие проектов ИИ в медицинской деятельности. В 2024 году объём финансирования в области медицинского искусственного интеллекта был оценен в 12.6 млрд. долл. Ожидается, что к 2033 году объём финансирования вырастет до 107.3 млрд. долл., при среднегодовом темпе роста 32.5% за период 2026–2033 г.¹

¹Healthcare Artificial Intelligence Market Analysis. URL: <https://www.marketresearchintellect.com/ru/product/healthcare-artificial-intelligence-market-size-and-forecast/> (Дата обращения: 14.11.2025)

Мировым трендом в медицине становится дистанционное общение с пациентом. В медицинских учреждениях Китая функционируют полностью «облачные клиники», как, например, Ping An Good Doctor, где диагностика и мониторинг заболеваемости пациента происходит в онлайн-формате.

Необходимость применения искусственного интеллекта в медицине обусловлена, в первую очередь, быстрым старением населения земного шара, ростом хронических и онкологических заболеваний у категории населения 60+, необходимостью обработки и анализа больших объёмов информации, постоянным увеличением расходов на здравоохранение и дефицитом квалифицированного медицинского персонала.

Несмотря на указанные глобальные процессы, применение ИИ в медицине сопряжено со значительными трудностями. Ниже выделим некоторые наиболее актуальные проблемы при использовании искусственного интеллекта в медицине:

- корректность постановки диагнозов и выдачи медицинских рекомендаций, требуется клиническая валидация на разных популяциях;
- проблема «черного ящика»: обучение модели происходит неизвестным субъектом (коллективом) и на неизвестном наборе данных;
- точность диагностирования;
- проблемы определения ответственности за ошибочные и ненадлежащие действия системы медицинского ИИ;
- определение состава конечного ИТ-продукта и его лицензирование;
- проблемы правовой регламентации систем искусственного интеллекта, в том числе, осуществляющих трансграничную медицинскую деятельность [2].

Решение указанных проблем требует подробного анализа используемых в настоящее время наборов данных и методов тестирования.

В Российской Федерации применение систем ИИ медицинского назначения регламентируется национальным стандартом РФ².

МЕДИЦИНСКИЕ БЕНЧМАРКИ, ЭТАПЫ И МЕТОДЫ ТЕСТИРОВАНИЯ

Для оценки качества технической составляющей медицинского ИИ используют эталонные наборы данных, так называемые бенчмарки (от англ. *benchmark* — «ориентир», «эталон»). Отметим, что в части аппаратного ресурса бенчмарк — это задача,

служащая эталонным тестом производительности компьютерной системы.

На эталонном наборе данных от системы медицинского ИИ ожидается получение диагноза и рекомендаций, совпадающих с общепринятыми в медицинской практике.

В настоящее время рассматриваются шесть популярных мультимодальных медицинских бенчмарков:

- **NEJM** (New England Journal of Medicine)
- **JAMA** (Journal of the American Medical Association)
- **VQA-RAD** (Radiology VQA dataset)
- **PathVQA** (Pathology VQA dataset)
- **SLAKE** (Structured medical VQA)
- **OmniMedVQA** (универсальный мультимодальный набор для медицины)

Большие языковые модели (БЯМ), такие как GPT-5, сегодня достигают высоких результатов на перечисленных медицинских бенчмарках [3]. При этом ведущие системы часто угадывают правильный ответ даже тогда, когда ключевые входные данные (например, изображения) удалены, меняют решения при незначительных изменениях формулировки запроса и создают убедительные, но ошибочные объяснения для своих выводов. Такие факты показывают, что современные бенчмарки удовлетворяют методикам тестирования, но не медицинскому пониманию процесса диагностики.

В статье [3] тестирование шести флагманских моделей медицинских ИИ на шести широко используемых бенчмарках обнаружило, что за высокими позициями в таблицах лидеров скрывается ненадежность («хрупкость») оценок.

Кроме того, необходимо отметить, что для постановки диагноза можно ввести ошибки первого рода (когда исходя из известного тестового набора симптомов поставлен неверный диагноз) и второго рода (когда заданный тест, содержащий признаки заболевания, системой не отнесен к содержащим заболевание).

Таким образом, оценка на медицинских бенчмарках не отражают напрямую готовность ИИ к реальному применению. Чтобы ИИ заслужил доверие в здравоохранении, нужно требовать большего, чем хорошие результаты на тестах — необходима проверка на устойчивость, корректные рассуждения и соответствие реальным медицинским требованиям. Устойчивость в данном случае — принятие одного и того же решения при небольших изменениях входных данных для анализа. Когда мы гово-

²Системы искусственного интеллекта в клинической медицине. Основные положения. ГОСТ Р 59921.0—2022.

рим о «рассуждении» модели, мы подразумеваем процесс формирования выводов по диагнозу в соответствии с запросом врача, а не самостоятельный субъектный процесс мышления, не свойственный языковой модели.

Напомним, что тестирование (испытание, исследование) программного обеспечения – процесс, имеющий своей целью проверку соответствия между действительным поведением программы и её ожидаемым поведением на конечном наборе испытаний (тестов), выбранных определённым образом (ISO/IEC TR 19759:2005)³.

Другими словами, процесс тестирования программного обеспечения – это структурированная последовательность действий, нацеленных на обеспечение качества продукта.

Заметим, что в идеальном виде **тестирование ПО включает в себя следующие этапы:**

- Анализ требований. Изучение документации и понимание бизнес-логики продукта.
- Планирование тестирования. Разработка стратегии и плана тестирования.
- Разработка тестовых сценариев. Создание тест-кейсов и чек-листов. Руководящие документы регуляторов РФ вводят понятие программ и методик тестирования.
- Подготовка тестовой среды. Настройка необходимого окружения для проведения тестов.
- Выполнение тестов. Непосредственная проверка функциональности продукта.
- Регистрация и классификация дефектов. Документирование найденных ошибок.
- Повторное тестирование и регрессионное тестирование. Проверка исправлений.
- Подготовка отчётов. Анализ результатов и составление итоговой документации.

Ниже приведем **некоторые методы тестирования ПО.**

Метод «чёрного ящика» предполагает, что тестирование проводится без глубоких знаний о внутренней работе ПО. Специалист не принимает во внимание архитектуру системы и не имеет доступа к исходному коду.

Согласно **методу «белого ящика»** эксперт проводит подробное исследование внутренней логики и структуры кода, изучает его изнутри и выясняет, какое устройство или блок кода ведёт себя некорректно.

Метод «серого ящика» означает, что специалист имеет представление о внутреннем устройстве ПО, но не слишком глубокое. Он ставит себя на место конечного пользователя, но проверяет функционал программы, опираясь на понимание её внутреннего устройства.

Можно констатировать, что тестирование с помощью бенчмарков относится к «серому тестированию», поскольку у тестировщика имеется только весьма общее представление о свойствах и устройстве большой медицинской модели.

ИЛЛЮЗИЯ ГОТОВНОСТИ МЕДИЦИНСКИХ СИСТЕМ ИИ

Как указано ранее (исследование [3]), высокие позиции в таблицах лидеров скрывают ненадежность оценок и обучение «по ярлыкам». С помощью экспертной оценки по клинической рубрике показано, что разные бенчмарки на самом деле измеряют разные качества, но при этом используются как взаимозаменяемые. Это маскирует сбои и искажает реальную готовность систем.

Кроме того, существующие системы медицинских ИИ отличаются наличием упомянутой выше проблемы доверия. Модели класса GPT-5 [4] вполне способны сдавать медицинские экзамены и лидировать в обработке мультимодальных бенчмарков (то есть, содержащих смешанные наборы данных, как правило, тест и изображение), но при этом они всё ещё допускают ошибки в простых задачах — например, не дают устойчивых и одинаковых ответов при перестановке вариантов, не могут обосновать диагноз медицински корректными рассуждениями. Это означает, что в целом тестирование бенчмарками некорректно. Поэтому недостаточно просто немного дорабатывать системы ИИ, необходим более фундаментальный подход [5].

Текущие медицинские бенчмарки больше напоминают материал для сдачи тестов, подобных задачам ЕГЭ, а не систематизированное медицинское знание [5]. Например, модель может условно «диагностировать» пневмонию не по анализу рентгенологических признаков, а потому, что симптомы «кашель» и «температура» статистически часто встречались вместе с диагнозом «пневмония» в обучающих данных. Это не является медицинским

³SWEBOK (Software Engineering Body of Knowledge) — международный стандарт ISO/IEC TR 19759 от 2015 г., в котором описана общепринятая сумма знаний по программной инженерии.

Документ был создан при сотрудничестве нескольких профессиональных организаций и предприятий и опубликован обществом IEEE Computer Society (IEEE)[2]. В 2005 году он был принят как стандарт ISO/IEC TR 19759:2005. В конце 2013 года была одобрена и опубликована новая версия SWEBOK V3, которая стала стандартом ISO/IEC TR 19759:2015. В 2016 году общество IEEE Computer Society создало комитет SWEBOK Evolution.

пониманием симптомов, скорее это можно назвать «обучением по ярлыкам».

При более тщательном анализе можно констатировать следующие **важные для тестирования факторы**:

- **Современные критерии успешности модели некорректны.** На мультимодальных (содержащих смешанные данные) медицинских бенчмарках [6, 7] ведущие модели сохраняли почти ту же точность даже при удалении изображений. Ответы на вопросы, явно требующие визуального ввода, они всё равно угадывали, при этом достаточно было одного знакомого по предварительному обучению варианта.

- **Неустойчивость работы.** Перестановка вариантов вводимых данных, «отвлекающие ответы» или незначительное изменение изображения приводили к резким сдвигам в предсказаниях, хотя сама медицинская суть задачи не менялась.

- **Сфабрикованные рассуждения.** Модели, обученные «думать шаг за шагом» [8], часто выдавали уверенные объяснения с некорректной логикой: либо строили медицински правдоподобное обоснование неправильного ответа, либо поддерживали правильный ответ вымышленным рассуждением (ошибки первого и второго рода).

Данные факты демонстрируют фундаментальные проблемы в оценке прогресса медицинского ИИ. При помощи упомянутых бенчмарков проверяется соответствие шаблону, а не «понимание» искусственным интеллектом сути медицинской проблемы. Они стабильны в формате тестов, но не эффективны в реальных условиях.

Медицинские решения в реальном мире принимаются в условиях неопределённости, неполноты информации и высокого риска в борьбе за жизнь и здоровье пациента. Если модель ведет себя некорректно при перестановке вариантов ответов, то можно ли доверять ей при неоднозначных симптомах или изображениях (например, рентгенологических снимках) плохого качества?

Таким образом, бенчмарки служат не только оценке моделей, но и формированию процесса их обучения и оптимизации [9]. Когда они поддерживают принятие решения «по ярлыкам», а не подлинное понимание, они формируют несоответствующую действительности оценку прогресса в медицине.

В статье [3] выявлены уязвимости с помощью серии **стресс-тестов** на шести передовых моделях и шести популярных мультимодальных медицинских бенчмарках. В ней показано, что хорошие и даже рекордные результаты скрывают неустойчивость работы, а успехи моделей часто зависят от особенностей конкретного набора, а не от их возможностей. Стесс-тест включает изменённые или

искажённые данные, создавая «стрессовую ментальную нагрузку» на медицинский ИИ.

Далее в [3] проанализированы бенчмарки. Помощью структурированного анализа под руководством врачей оценивалось соответствие требованиям к сложности рассуждений и зависимости от изображения. Результаты оказались показательными: одни и те же популярные бенчмарки тестируют разные качества, но при этом воспринимаются как равнозначные. Это скрывает сбои и искажает реальную готовность моделей к внедрению. На рисунках 1-4 отражена выявляемая стресс-тестами скрытая неустойчивость мультимодальных моделей в медицинских задачах.

На рис. 2 видно, что традиционные оценки (верхняя линия) показывают прогресс, но показатели устойчивости (нижняя линия) демонстрируют растущую ненадежность при неполных или «враждебных» входных данных.

Результаты стресс-тестов, отраженные на рис. 3, разрушают ошибочное впечатление от таблицы лидеров, показывая специфические для каждой модели сбои.

На рис. 4 проиллюстрированы примеры правильного ответа даже без изображения и при перемешанных вариантах (обучение по ярлыкам), а также сконструированное рассуждение при ответе без визуальных данных. Ответы сгенерированы GPT-5 при запросе объяснения «постфактум». Возможны вариации ответа при разных запусках. Чтобы заставить модели отвечать напрямую, во всех промптах использовался запрос «пожалуйста, укажите только индекс и содержание варианта».



Рис.1. Концептуальная иллюстрация скрытых уязвимостей моделей

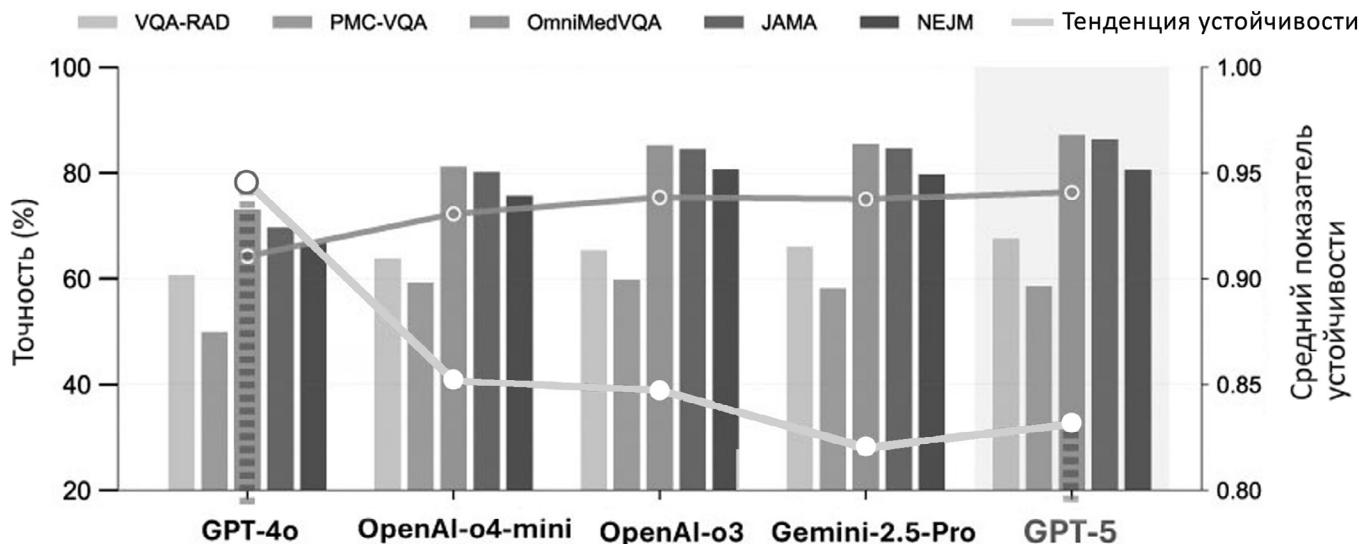


Рис.2. Традиционные оценки и тенденция устойчивости

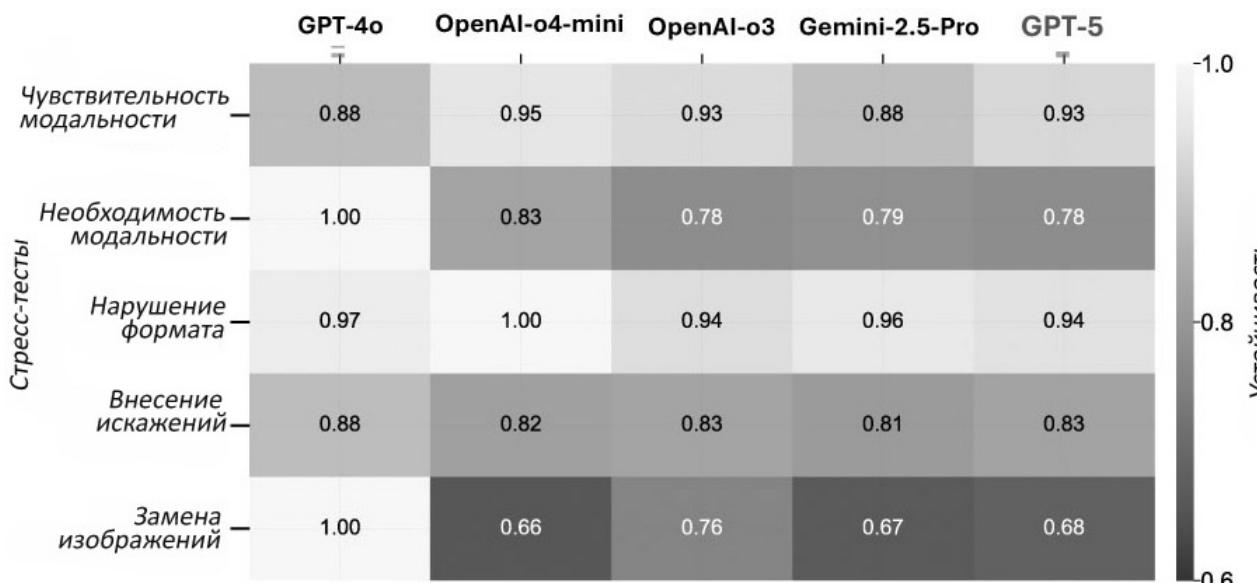


Рис.3. Показатели устойчивости по проведенным стресс-тестам

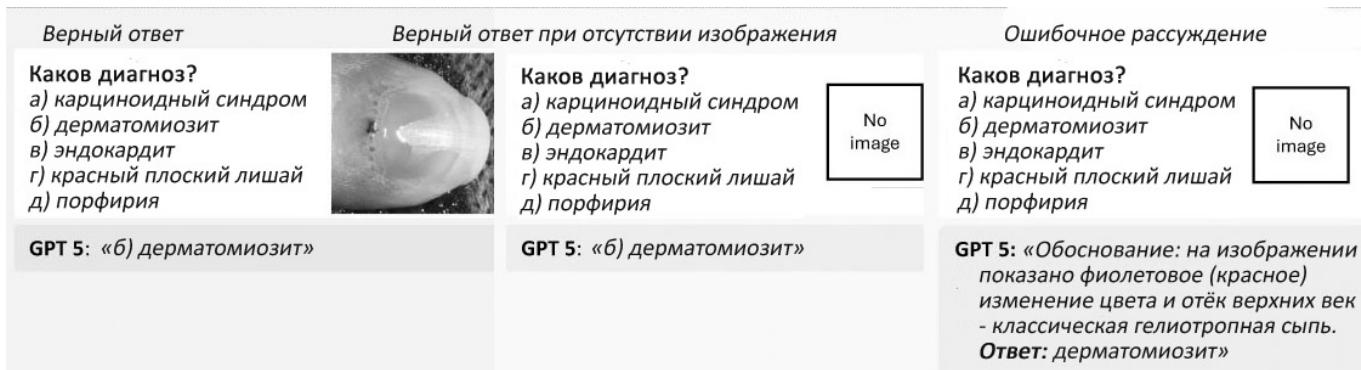


Рис.4. Примеры нарушения корректности рассуждений

Приводимые на рисунках 1-4 данные и примеры показывают необходимость радикального пересмотра методов оценки эффективности медицинского ИИ. Прежде чем такие системы будут использоваться для поддержки принятия медицинских решений, важно убедиться, что они достигают правильного ответа исходя из логики, а не просто потому, что умеют сдавать тест.

Для этого в [3] предложена серия целевых стресс-тестов, устраниющих лишние подсказки и возможности срабатывания «по ярлыкам». Даже в этих жёстких условиях видно улучшение поколений моделей, что указывает на рост реальных возможностей. Однако этот рост остаётся ограниченным, а неустойчивость сохраняется.

Дальнейший прогресс требует улучшения не только самих моделей, но и способов их оценки.

СТРЕСС-ТЕСТИРОВАНИЕ МОДЕЛЕЙ НА МУЛЬТИМОДАЛЬНЫХ МЕДИЦИНСКИХ БЕНЧМАРКАХ

Высокие результаты на бенчмарках могут создавать видимость устойчивости, но на деле скрывают ненадежность при пропущенных данных или лёгких искажениях. В [3] предлагается набор стресс-тестов, которые имитируют неопределённость и показывают поведение моделей в условиях, приближённых к реальной клинической практике. Вместо того, чтобы оценивать только правильность ответа, эти тесты раскрывают, как модели реагируют на неполный, искажённый или «враждебный» ввод.

Стресс-тесты целесообразно распределить по уровням сложности: от недостающих или вводящих в заблуждение модальностей до изменения структуры и качественного анализа рассуждений.

Устойчивость к модальностям

Стресс-тест 1: чувствительность к модальностям

- **Цель:** проверить, как модели реагируют на удаление изображений в мультимодальных медицинских бенчмарках.

- **Настройка:** оценивалось шесть моделей (GPT-5, Gemini-2.5 Pro [10], OpenAI-o3 [11], OpenAI-o4-mini [11], GPT-4o [12] и DeepSeek-VL2 [13]) на бенчмарках NEJM [6] и JAMA [7] в двух режимах: «изображение и текст» (по умолчанию) и «только текст». Все задания — это вопросы с несколькими вариантами, основанные на клинических случаях с изображениями (радиология, дерматология, патология и др.).

• Результаты

Удаление изображений показало резкие различия в поведении моделей, которые в обычных усло-

виях демонстрируют схожие результаты.

NEJM:

- о GPT-5: 80,89% → 67,56% (падение на 13,33 п.п.).
- о Gemini-2.5 Pro: 79,95% → 65,01% (-14,94 п.п.).
- о OpenAI-o3: 80,89% → 67,03% (-13,86 п.п.).
- о GPT-4o: 66,90% → 37,28% (-29,62 п.п.).
- о OpenAI-o4-mini: 75,91% → 66,49% (-9,42 п.п.).
- о DeepSeek-VL2: 33,16% → 25,30% (очень низкие показатели в обоих режимах).

JAMA:

- о GPT-5: 86,59% → 82,91% (-3,68 п.п.).
- о Gemini-2.5 Pro: 84,84% → 74,93% (-9,91 п.п.).
- о OpenAI-o3: 84,75% → 82,65% (-2,10 п.п.).
- о OpenAI-o4-mini: 80,50% → 78,40% (-2,10 п.п.).
- о GPT-4o: 69,90% → 68,90% (-1,00 п.п.).
- о DeepSeek-VL2: 38,20% → 32,60%.

Можно сделать вывод о том, что модели с похожими результатами при полном вводе (текст и изображение) резко расходятся по результатам без изображений. Таким образом, демонстрируется скрытая зависимость от модальности, не отражённая в рейтингах. Сами бенчмарки различаются по «визуальной зависимости»: на NEJM удаление изображений резко снижает точность, а на JAMA — почти нет (многие задачи решаются только по тексту).

Как видно на рис. 5, точность падает при удалении изображений из двух диагностических бенчмарков. Резкое падение на NEJM, но минимальное — на JAMA говорит о непоследовательной визуальной зависимости в разных наборах и типах вопросов. Светлые сегменты показывают точность без изображений (только текст), тёмные — уровень снижения точности при переходе от данных, включающих изображение и текст, к данным, содержащим только текст.

Эти различия мотивируют перейти к профилированию заданий на уровне отдельных элементов.

Стресс-тест 2: необходимость в модальности

- **Цель:** проверить, могут ли модели отвечать на вопросы, где для правильного решения нужны и текст, и изображение.

- **Настройка:** сформировано подмножество из 175 заданий NEJM, которые, по мнению врачей, невозможно решить без изображения. Из текстов были убраны все подсказки, чтобы правильный ответ требовал именно визуального анализа. Протестированы те же шесть моделей (GPT-5, Gemini-2.5 Pro, OpenAI-o3, OpenAI-o4-mini, GPT-4o и DeepSeek-VL2) в двух режимах: «изображение и текст» и «только текст».

• Результаты

В режиме «изображение и текст»:

- о GPT-5: 66,28%

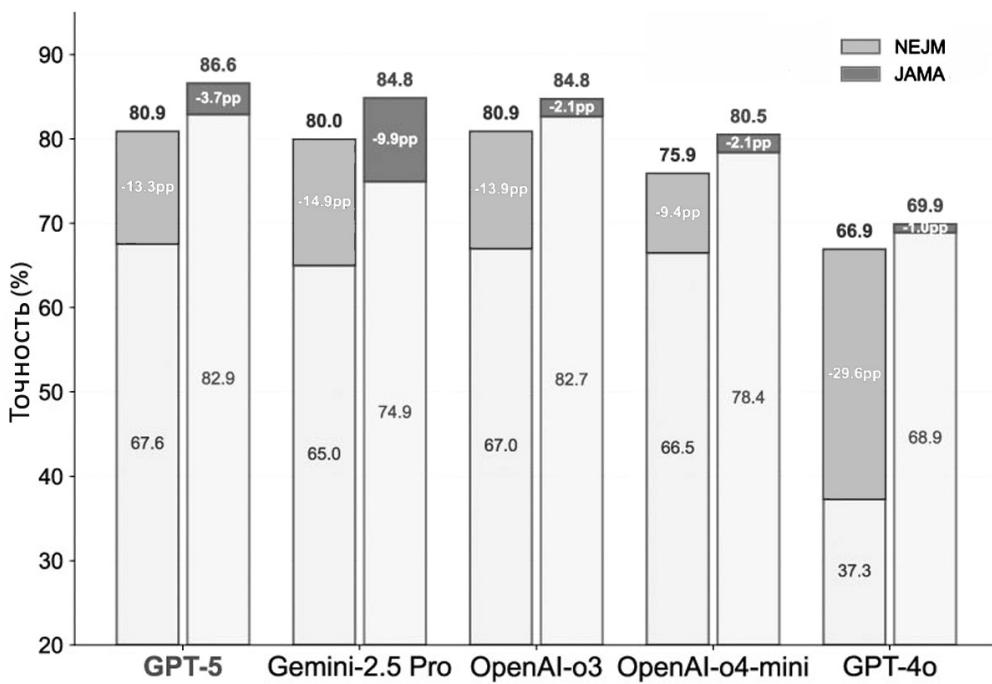


Рис. 5. Стресстесты на устойчивость к модальностям (на бенчмарках NEJM и JAMA)

- Gemini-2.5 Pro: 67,42%
- OpenAI-o3: 61,71%

Все три модели показали, что задания решаемы при полном вводе.

В режиме **«только текст»**:

- GPT-5: 37,71%
- Gemini-2.5 Pro: 37,14%
- OpenAI-o3: 37,71%
- OpenAI-o4-mini: 33,71%
- GPT-4o: всего 3,4% (из-за отказа отвечать без изображения).

Модели показывают результат выше случайного уровня (20%), даже когда изображения удалены, хотя они необходимы для ответа. Это указывает на использование артефактов обучающего набора данных (датасета): частотных шаблонов, статистических ассоциаций или заученных пар «вопрос – ответ».

GPT-4o выделяется: она часто отказывается отвечать без изображения. Хотя это снижает точность численно, такой подход отражает более осторожное поведение — модель сигнализирует об отсутствии данных вместо того, чтобы угадывать.

Важно отметить, что такое поведение не обязательно говорит о «плохой практике» модели. Многие примеры из бенчмарков доступны публично, и модели могли непреднамеренно запомнить их во время предобучения.

Однако более общий вывод заключается в том, что высокие результаты на бенчмарках не всегда отражают подлинное мультимодальное понимание. Часто они возникают из-за артефактов набора данных или заученных ассоциаций. Это создаёт ложное

ощущение готовности модели для реального медицинского применения.

Результаты на 175 заданиях NEJM, где ответ требует изображения, проиллюстрированы на рис. 6. Большинство моделей показывают результат выше случайного (20%) даже без изображений. Это свидетельствует о том, что модель опирается на ярлыки: частотные шаблоны, статистические ассоциации или запомненные пары «вопрос–ответ». GPT-4o при этом даёт результаты ниже случайного уровня, так как часто отказывается отвечать при отсутствии изображения.

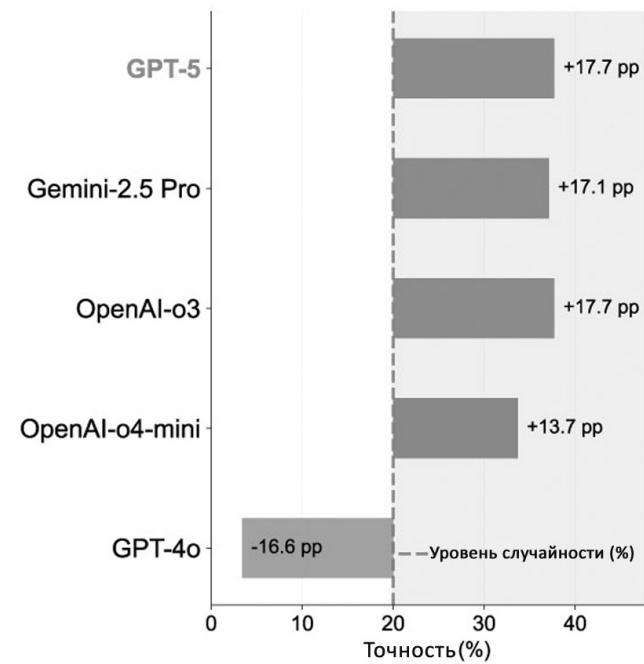


Рис.6. Необходимость в модальности

Нарушение паттернов-ярлыков

Стресс-тесты 1–2 показали, что при отсутствии одной из модальностей мультимодальные модели нередко прибегают к ярлыкам. Более детально их изучить возможно посредством целенаправленного введения искажений во входные данные.

Стресс-тест 3: нарушение структуры

● **Цель:** проверить, опираются ли модели на представление вариантов ответов (например, фиксированные позиции) вместо содержания вопроса.

● **Настройка:** использовались те же 175 заданий NEJM, где для генерации правильного ответа необходимы и текст, и изображение. Распределение правильных ответов было равномерным. Варианты ответа были случайно переставлены (с сохранением смысла). Верный вариант оставался тем же.

● Результаты

Без изображений (только текст): точность падала при перестановке ответов, например:

- GPT-5: 37,71% → 32,00%
- Gemini-2.5 Pro: 37,14% → 33,14%
- OpenAI-o3: 37,71% → 31,42%

С изображениями (изображение и текст): в большинстве случаев точность немного росла, например:

- GPT-5: 66,28% → 70,85%
- Gemini-2.5 Pro: 67,42% → 70,28%
- OpenAI-o3: 61,71% → 64,00%

Таким образом, в режиме «только текст» модели зависят от поверхностных шаблонов (например, позиции ответа). Перестановка «ломает» этот ярлык и точность падает.

В режиме «изображение и текст» ситуация стабильнее: визуальные данные помогают компенсировать текстовые ярлыки. Данный вывод подтверждает, что с дополнительными модальностями результаты меньше искажаются (рис. 7).

Стресс-тест 4: замена отвлекающих ответов

● **Цель:** проверить, используют ли модели заученные пары «вопрос–вариант» без реального понимания.

● **Настройка:** постепенно заменяли от 1 до 4 отвлекающих вариантов случайными ответами из других заданий того же набора. Правильный вариант сохранялся. Также тестировался вариант, где один отвлекающий ответ заменялся словом «Unknown» («Неизвестно»). Проверены режимы «только текст» и «изображение и текст».

● Результаты

Только текст:

○ Чем больше ответов заменяли, тем хуже был

результат → модели приближались к случайному угадыванию.

○ GPT-5: 37,7% → 20,0% (при замене всех 4 отвлекающих ответов).

Изображение и текст:

○ Точность, наоборот, возрастала при замене отвлекающих.

○ GPT-5: 66,3% → 90,9% (при замене всех 4).

Замена словом «неизвестно»:

○ Незначительно возрастала точность у всех моделей (+5–8 п.п.).

○ GPT-5: рост примерно на +5,25 п.п.

В режиме «только текст» модели сильно зависят от знакомых паттернов между вопросами и вариантами. Когда эти паттерны нарушены, точность снижается. Добавление варианта «неизвестно» не воспринимается как сигнал «отказа от ответа», а как «слабый отвлекающий». В итоге модели становятся увереннее в выборе правильного ответа.

В режиме «изображение и текст» – наоборот: чем меньше ярлыков в отвлекающих, тем проще задачу решать визуально (рис. 7).

Таким образом, то, что выглядит как «мультимодальная компетентность», на деле может быть поверхностной стратегией исключения, при этом небольшие изменения в структуре задачи ведут к неустойчивости поведения.

Стресс-тест 5: замена изображения

● **Цель:** проверить, действительно ли модели интегрируют визуальные и текстовые данные, или опираются на поверхностные ассоциации «изображение–ответ».

● **Настройка:** было подобрано 40 заданий из NEJM, где правильный ответ жёстко связан с конкретными визуальными признаками. Затем исходное изображение заменялось другим – соответствующим отвлекающему варианту. Вопрос и текстовые варианты остались без изменений. В новой версии правильным становится отвлекающий вариант.

● Результаты

○ GPT-5: 83,3% → 51,7% (падение на 31,6 п.п.).

○ Gemini-2.5 Pro: 80,8% → 47,5% (-33,3 п.п.).

○ OpenAI-o3: 76,7% → 52,5% (-24,2 п.п.).

○ OpenAI-o4-mini: 71,7% → 37,5% (-34,2 п.п.).

○ GPT-4o: практически без изменений (36,7% → 41,7%).

Происходит резкое падение точности при минимальном изменении входных данных, что показывает: многие модели опираются на «заученные пары» (определенный визуальный паттерн → определённый ответ), а не на полноценное сопоставление изображения и текста. Исключением

оказался GPT-4o, который почти не изменил результат (рис. 7).

Хотя сложность задач могла меняться (новые изображения могли быть более неоднозначными), консистентное падение точности у топовых моде-

лей подтверждает, что высокие бенчмарк-баллы часто связаны с ярлыками, а не с интеллектуальной и устойчивой визуально-текстовой обработкой данных.

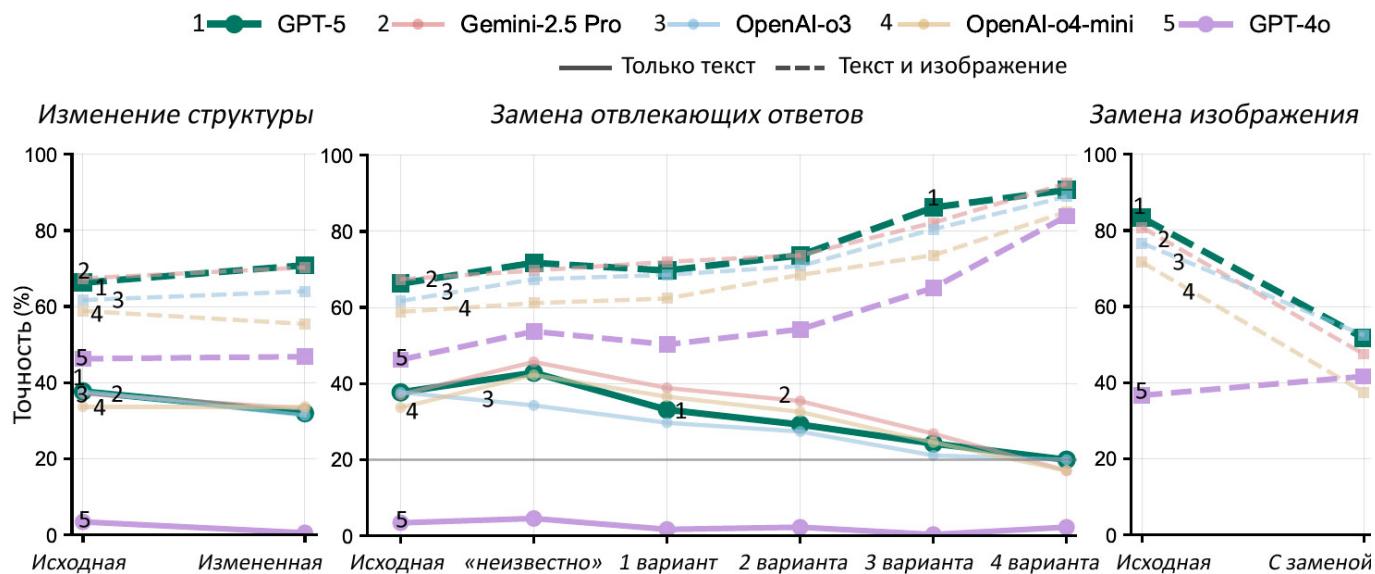


Рис.7. Результаты моделей при изменении структуры тестов, замене отвлекающих ответов и изображений

Сигналы рассуждений

Стресс-тест 6: Качество рассуждений

• **Цель:** оценить, насколько рассуждения моделей на мультимодальных медицинских задачах соответствуют реальному пониманию.

• Настройка:

1. Использован способ, при котором модель показывает промежуточные шаги рассуждения (Chain-of-Thought, CoT) [14] для ~100 примеров из NEJM и VQA-RAD [15], чтобы проверить, улучшает ли пошаговое рассуждение результат.

2. Объяснения моделей проверены «вручную» с учетом оценки их фактичности, визуальной опоры и соответствия финальному ответу.

3. На датасете OmniMedVQA [16] варьировалась «сила рассуждений» (от слабых к сильным) у модели OpenAI-o3, для оценки качества рассуждений.

• Результаты

Chain-of-Thought: на VQA-RAD и NEJM при использовании способа CoT последовательно снижалась точность у большинства моделей. Исключение — OpenAI-o4-mini (+2–5 п.п.), вероятно потому, что модели меньшего размера выигрывают от «подсказки рассуждать пошагово». На бенчмарке OmniMedVQA усиление рассуждений давало лишь небольшие улучшения.

Проверка рассуждений вручную выявила следующие особенности работы моделей:

1. Уверенная, но ложная эвристика. Модель применяет правдоподобное, но неверное правило (например, выводит диагноз из частых ассоциаций), что приводит к систематическим ошибкам (рис. 8).

2. Правильный ответ, в основе которого — «галлюцинирование» при рассуждениях. Модель выбирает верный вариант, хотя не имела доступа к изображению, т.е. решение основано на ярлыках, а не на визуальном анализе (рис. 8).

3. Рассуждение при наличии изображения, но ошибочное. Модель разбивает решение на шаги, якобы анализируя изображение, но опирается на неверные или вымышленные детали. Итог: уверенный, но неправильный вывод (рис. 9).

Таким образом, были выявлены правильные ответы с ложной логикой (например, описание несуществующих признаков), усиленное визуальное непонимание (если модель ошибалась в визуальной интерпретации, то дальнейшие рассуждения лишь укрепляли неправильный вывод) и выполнение моделями интерпретируемых, но ложных шагов (модель делала пошаговое описание, но оно не вело к правильному ответу).

Из рассмотренных результатов стресс-тестов можно сделать вывод, что рассуждения моделей часто не связаны с настоящим мультимодальным пониманием. Правильный ответ может быть результатом подбора шаблона, а не анализа, в то врем-

мя как обоснования могут звучать убедительно, но включать фактические ошибки или вымышленные детали. Даже структурированные рассуждения часто не помогают найти верный ответ.

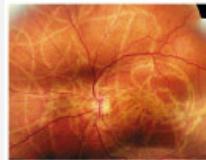
Это подчёркивает ненадёжность «сигналов рассуждений» и указывает на необходимость глубже изучать требования к рассуждению на базе медицинских бенчмарков.

Данные стресс-тесты показали, что высокие результаты могут скрывать неустойчивость. Но остаётся вопрос: что именно измеряют сами бенчмарки? Чтобы это выяснить, был проведен структурирован-

Уверенная, но ложная эвристика

Каков диагноз? Выберите один из вариантов ответа ниже.

- А) Ангиоидные полосы
- Б) Глаукома
- В) Макулярная дегенерация
- Г) Офтальмомиаз (эталонные данные)
- Д) Эластическая псевдоксантома



Рассуждение

На фотографии глазного дна видны множественные неровные, темные, похожие на трещины линии, расходящиеся от диска зрительного нерва проходящие под сосудами сетчатки. Эта классическая картина ангиоидных полос — разрывов кальцинированной/хрупкой мембранны Бруха. В макулярной области наблюдаются рубцевание/пигментные изменения, характерные для осложнений ангиоидных полос (например, хориоидальная неоваскуляризация и субретинальный фиброз), которые могут привести к потере зрения.

Почему другие варианты маловероятны

- Глаукома — будет видна экскавация диска зрительного нерва, а не перипапиллярные радиальные полосы.
- Макулярная дегенерация — друзы/географическая атрофия с центром в макуле, а не радиальные трещины.
- Офтальмомиаз — будут видны подвижные личинки, а не фиксированные полосы.
- Эластическая псевдоксантома — системное заболевание, часто связанное с ангиоидными полосами.

Ответ

А: Ангиоидные полосы

ный анализ (профилирование) заданий с участием врачей [3].

МЕТОДИКА ПРОФИЛИРОВАНИЯ

Использованы шесть популярных мультимодальных медицинских бенчмарков:

- **NEJM** (New England Journal of Medicine)
- **JAMA** (Journal of the American Medical Association)
- **VQA-RAD** (Radiology VQA dataset)

Правильный ответ, в основе которого — галлюцинирование при рассуждениях

Вопрос

Каков диагноз? Выберите один из вариантов ответа ниже.

- А) Карциоидный синдром
- Б) Дерматомиозит (эталонные данные)
- В) Эндокардит
- Г) Красный плоский лишай
- Д) Порфирия

No image

Рассуждение

На изображении показана фиолетовое (красное) изменение цвета ткани и отёк верхних век — классическая «гелиотропная сыпь». Эти признаки характерны для дерматомиозита. Почему другие варианты маловероятны:

- Карциоидный синдром: приливы, а не персистирующая фиолетовая периорбитальная сыпь.
- Эндокардит: периферические стигматы (узелки Ослера, поражения Джейнвея), а не гелиотропная сыпь.
- Красный плоский лишай: зудящие, полигональные, пурпурные папулы на запястьях/лодыжках со стриями Уикхема; не периорбитальное расположение.
- Порфирия: волдыри, светочувствительность на тыльной стороне кистей и гипертрихоз, а негелиотропная сыпь на веках.

Ответ:

Б) Дерматомиозит

Рис. 8. Ложная эвристика и «галлюцинирование» при рассуждениях

Рассуждение при наличии изображения, но ошибочное

Вопрос
Есть ли расширение желчных протоков?
Предоставьте анализ и обоснование ответа.

Рассуждение:

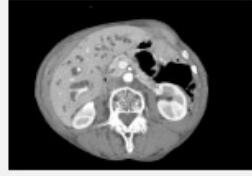
1) Вычислить базовые аналитические данные по изображению (например, размер), обрезать и увеличить часть изображения.

2) Проанализировать увеличенное изображение и перефокусировать: хм, похоже, это снимок нижней части брюшной полости, рядом с поджелудочной железой. Печень не отображается, и я вижу обе почки, которые выглядят симметричными. Желчный пузырь также не виден — не уверен насчёт общего желчного протока. Мне нужно будет внимательно проверить; похоже, расширения протоков нет, так что всё выглядит нормально... поскольку печень нечётко видна — возможно, отсутствует только левая доля. Я заметил расширенную структуру рядом с воротной веной, возможно, общий желчный проток. Рядом с поджелудочной железой также есть подпечёночная трубка, поэтому мне нужно будет проверить более широкую область.

3) Анализ и заключение

Хорошо, сейчас меньше внимания уделяю почкам. На этом срезе я замечаю две округлые области низкой плотности рядом с поджелудочной железой и воротами печени — одна из них может быть общим желчным протоком. Между двенадцатиперстной кишкой и общим желчным протоком есть удлинённая трубчатая структура, не слишком большая. Похоже, значительного расширения нет. При панкреатите типичный вопрос — о расширении желчных протоков, но здесь его нет — они нормального диаметра, без «симптома двойного протока».

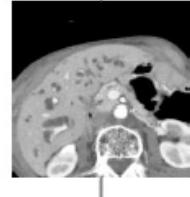
Ответ:
Нет. Расширения желчных протоков нет.



Обрезать и увеличить



Проанализировать, увеличить



Проанализировать

Рис. 9. Рассуждение при наличии изображения, но ошибочное

- **PathVQA** (Pathology VQA dataset)
- **SLAKE** (Structured medical VQA)
- **OmniMedVQA** (универсальный мультимодальный набор для медицины)

Ниже приведем критерии оценки заданий экспертами.

Требуемая сложность рассуждения:

- низкая: прямое извлечение факта или простой поиск;
- средняя: требуется сопоставление данных или базовая логика;

- высокая: комплексные рассуждения, интеграция нескольких шагов, клиническое мышление.

Зависимость от изображения:

- низкая: ответ можно дать по тексту;
- средняя: изображение помогает, но ответ можно дать и без него;
- высокая: без изображения ответить невозможно.

Результаты профилирования проиллюстрированы при помощи «карты профилей» на рис. 10.

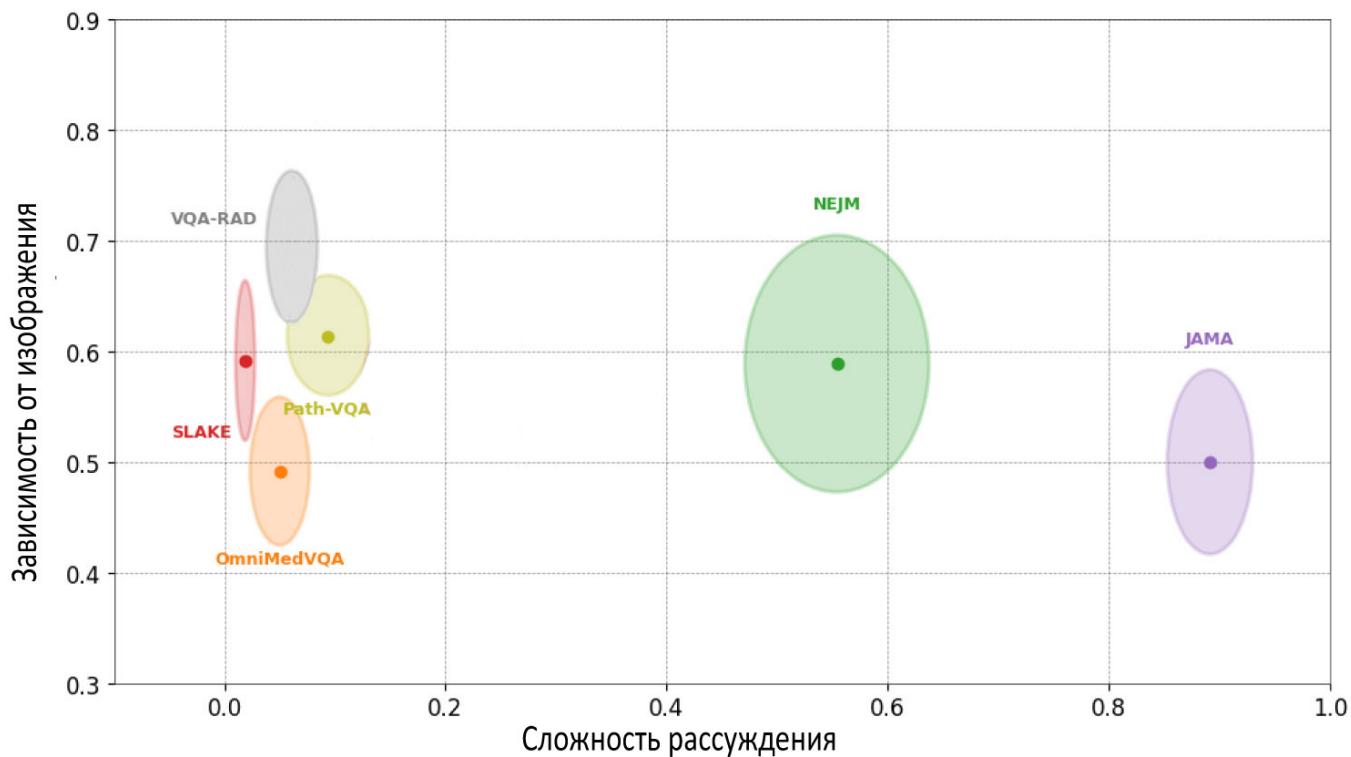


Рис.10. «Карта профилей» шести бенчмарков

- **NEJM**: высокий уровень сложности рассуждения, высокая зависимость от визуальных данных.
- **JAMA**: средняя сложность рассуждения, низкая зависимость от изображения.
- **VQA-RAD**: низкая сложность рассуждения, средняя зависимость.
- **PathVQA**: средняя сложность рассуждения, высокая зависимость от изображения.
- **SLAKE**: низкая сложность рассуждения, низкая зависимость от изображения.
- **OmniMedVQA**: широкий спектр сложностей и зависимостей, охватывающий разные сценарии.

Несмотря на то, что все эти наборы часто рассматриваются как взаимозаменяемые «тесты медицинской компетентности», на самом деле они измеряют разные аспекты.

• **NEJM** действительно проверяет интеграцию текста и изображений, требуя клинической логики.

• **JAMA** больше похож на текстовый экзамен с иллюстрациями — большинство заданий решаются без изображения.

• **VQA-RAD** в основном состоит из простых вопросов («где находится объект?» или «какая структура обозначена?»), требующих базового визуального восприятия, но не клинической логики.

• **PathVQA** ближе к NEJM по визуальной зависимости, но его вопросы чаще носят фактологический характер.

• **SLAKE** ограничен простыми задачами, часть из которых вообще решается по ключевым словам.

• **OmniMedVQA** наиболее разнообразен и охватывает широкий диапазон типов задач.

Это объясняет, почему модель может хорошо работать на одном бенчмарке и весьма неудовлетворительно — на другом, и почему суммарные рейтинги создают иллюзию универсальной готовности.

Из краткой характеристики, приведенной выше, видно, что бенчмарки действительно различаются по сложности и мультимодальности. Таким образом, их нельзя рассматривать как взаимозаменяемые тесты. Модель может казаться «лучшей» только потому, что она лучше справляется с лёгкими наборами, или наоборот: модель, сильная в сложных задачах, проигрывает в усреднённых рейтингах из-за более простых обучающих датасетов, поэтому сравнение моделей «по одной шкале» и усреднение результатов по разным наборам искажает картину.

Можно констатировать, что популярные лидер-борды создают иллюзию прогресса, скрывая при этом ненадежность моделей. Они создают впечатление поступательного движения в медицинском ИИ: каждая новая модель «поднимает планку» на нескольких бенчмарках. Однако проведенный по каждой модели анализ показывает, что эти цифры могут быть обманчивыми. Ниже приведем основные причины ненадежности показателей.

● **Разные бенчмарки — разные навыки.** Как было показано выше, NEJM формирует клиническое рассуждение, а JAMA — в основном текстовое распознавание. Когда результаты усредняются, создаётся иллюзия универсальной компетентности.

● **Рост показателей не гарантирует устойчивости.** Стресстесты выявили, что новые модели могут быть даже более ненадёжными, чем предшественники, несмотря на рост общей точности.

● **Ненадёжные обоснования.** Модели часто объясняют свои ответы убедительно, но ошибочно. Это усиливает иллюзию понимания.

Восприятие высоких результатов без контекста приводит к выводу, что модели уже готовы к использованию в медицине. Но на деле они могут, как показано выше,искажать ответ при изменениях структуры, подменять рассуждение шаблонами, демонстрировать успех на лёгких наборах и неудачу — на сложных.

УРОКИ СТРЕСС-ТЕСТОВ И ЗАДАЧИ СОВЕРШЕНСТВОВАНИЯ СИСТЕМ

Цитируемые исследования выявляют систематические уязвимости в мультимодальных медицинских моделях:

- опора на ярлыки вместо понимания;
- неустойчивость при изменениях структуры;
- обоснования, которые звучат убедительно, но вводят в заблуждение.

Эти факты имеют важные последствия как для оценки прогресса, так и для доверия к медицинскому ИИ.

В настоящее время высокие результаты на медицинских наборах часто трактуются как сигнал готовности к внедрению в здравоохранение. Но стресс-тесты показывают, что это ошибочное впечатление. В реальных клинических условиях входные данные часто неполные, противоречивые или искаженные. Если модель, в дополнение к неполным и противоречивым данным, дает сбой при перестановке вариантов ответа, становится очевидно, что ей нельзя доверять при работе с пациентами.

Кроме того, можно констатировать, что наличия одних медицинских бенчмарков недостаточно для того, чтобы медицинский ИИ заслужил доверие. Необходимо проводить тесты на устойчивость (способность модели работать при изменениях структуры, частичной потере данных), анализ на соответствие требованиям к прохождению таких тестов моделями и оценку качества рассуждений, а не только конечного ответа. Проверка клинической эффективности должна проводиться с привлечени-

ем консилиума врачей, а не только автоматических метрик.

Чтобы говорить о реальной готовности медицинского ИИ, нужно учитывать три уровня:

1. **Уровень бенчмарков.** Бенчмарки полезны, но ограничены, они выявляют лишь частные аспекты (например, знание терминов, способность сопоставлять изображение и текст).

2. **Уровень устойчивости модели.** Модель должна сохранять надёжность при сдвигах структуры, вариациях данных и отсутствии части информации.

3. **Уровень клинической практики.** Ключевой критерий — польза и безопасность для пациента. Это можно оценить только в тесном сотрудничестве с медицинскими экспертами и в условиях реальной практики.

Можно сформулировать следующие перспективные направления для совершенствования как самих систем на основе мультимодальных моделей, так и методов их оценки:

1. Автоматизация стресс-тестов для масштабируемой проверки.

2. Создание новых бенчмарков, более близких к реальным клиническим задачам, селектирование бенчмарков по отраслям клинической практики, обязательный учет возрастных, расовых и гендерных особенностей пациентов.

3. Совместная работа с врачами для определения метрик, которые действительно отражают надежность и эффективность диагностирования.

4. Учет предварительной клинической картины — добавление к текущим симптомам или результатам обследований пациента его истории болезни, данных из медицинской карты или предыдущих диагнозов, т.е. переход от методологии интеллекта к методологии сознания, когда выводы делаются на основе не только текущих данных, но и некоторой предыстории. [17]

ЗАКЛЮЧЕНИЕ

На современном этапе развития медицинских систем ИИ очевидно, что медицинские бенчмарки нельзя использовать как прямой показатель готовности к клиническому внедрению, поскольку они дают частичную картину, которая может быть полезной для исследования, но не отражает реальную надёжность модели.

В частности, исследование [3] выявило такие систематические уязвимости в мультимодальных медицинских моделях, как опора на ярлыки при отсутствии понимания, неустойчивость при небольших изменениях структуры, способность приводить

убедительные обоснования, которые вводят врача в заблуждение. Результаты данного исследования имеют важные последствия для оценки прогресса в области медицинского ИИ, а также для объективной оценки уровня доверия к нему.

Сегодня высокие результаты на медицинских бенчмарках часто трактуются как сигнал готовности к внедрению в здравоохранение. Но это является серьезным заблуждением, поскольку в реальных клинических условиях входные данные часто не обладают свойствами полноты, непротиворечивости и достоверности. Кроме того модели, меняющей результат даже при простой перестановке одинаковых вариантов ответа, нельзя доверять при работе с критичными медицинскими данными.

Чтобы медицинский ИИ заслужил доверие, необходимо систематически проводить:

- тесты на устойчивость, т.е. способность работать при изменениях структуры, частичной потере данных и т. д.;

- оценку качества рассуждений, а не только конечного ответа;

- проверку клинической полезности с привлечением врачей, а не только автоматических метрик.

Для объективной оценки готовности медицинского ИИ нужно учитывать три уровня: уровень бенчмарков, уровень устойчивости моделей и уровень клинической практики. Реальную пользу медицинского ИИ возможно оценить только в тесном сотрудничестве с медицинскими экспертами и лечащими врачами в условиях реальной клинической практики.

СПИСОК ЛИТЕРАТУРЫ

1. Антонова Н.В., Бальхаева С.Б., Гаунова Ж.А. и др. Юридическая концепция роботизации: монография // Ответственные редакторы Ю.А. Тихомиров, С.Б. Нанба. – Москва: Проспект, 2019. – С. 232.
2. Имамеева Р.Д. Риски создания и функционирования искусственного интеллекта в медицине // Вестник Московского университета имени С.Ю. Витте. Серия 2 «Юридические науки». № 1 (27) 2021. С. 33-40.
3. Yu Gu, Jingjing Fu, Xiaodong Liu, Jeya Maria Jose Valanarasu, et al. The Illusion of Readiness: Stress Testing Large Frontier Models on Multimodal Medical Benchmarks. Microsoft Research, Health & Life Sciences. 2025. URL: <https://arxiv.org/pdf/2509.18234> (Дата обращения: 25.09.2025)
4. OpenAI. GPT-5. 2025. URL: <https://openai.com/index/introducing-gpt-5/> (Дата обращения: 25.09.2025)
5. Rajpurkar P., et al. The Current and Future State of AI Interpretation of Medical Images. URL: <https://www.nejm.org/doi/full/10.1056/NEJMra2301725> (Дата обращения: 25.09.2025)
6. NEJM Image Challenge Dataset. URL: <https://www.nejm.org/image-challenge> (Дата обращения: 25.09.2025)
7. JAMA Clinical Challenge Dataset. URL: <https://jamanetwork.com/collections/44038/clinical-challenge> (Дата обращения: 25.09.2025)
8. Wei J., et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. URL: <https://arxiv.org/abs/2201.11903> (Дата обращения: 25.09.2025)
9. Sagawa S., et al. WILDS: A Benchmark of in-the-Wild Distribution Shifts. URL: <https://www.researchgate.net/scientific-contributions/Shiori-Sagawa-2162238023/publications/1#articles> (Дата обращения: 25.09.2025)
10. Gemini-2.5 Pro, Google DeepMind. URL: <https://deepmind.google/models/gemini/pro/> (Дата обращения: 25.09.2025)
11. OpenAI-o3 и OpenAI-o4-mini. URL: <https://openai.com/index/introducing-o3-and-o4-mini/> (Дата обращения: 25.09.2025)
12. OpenAI GPT-4o. URL: <https://openai.com/index/hello-gpt-4o/> (Дата обращения: 25.09.2025)
13. DeepSeek-VL2. URL: <https://github.com/deepseek-ai/DeepSeek-VL2?ysclid=mhxn5446zr649625886> (Дата обращения: 25.09.2025)
14. Kojima T., et al. Large language models are zero-shot reasoners. URL: https://www.researchgate.net/publication/360834082_Large_Language_Models_are_Zero-Shot_Reasoners. (Дата обращения: 25.09.2025)
15. Lau J., et al. A dataset of clinically generated visual questions and answers about radiology images. URL: <https://colab.ws/articles/10.1038%2Fdata.2018.251> (Дата обращения: 25.09.2025)
16. Yutao Hu, et al. OmniMedVQA: A New Large-Scale Comprehensive Evaluation Benchmark for Medical LVLM. URL: <https://arxiv.org/html/2402.09181v2> (Дата обращения: 25.09.2025)
17. Щербаков А.Ю. Практическая модель искусственного сознания // Вестник современных цифровых технологий. 2025. № 24. С. 4-12.

УДК: 004.9, 338.2

Создание цифрового двойника угледобывающего предприятия на основе геопространственных данных¹

T.V. Chaplygina, K.I. Peters, R.E. Zverev,
E.S. Gromov, S.A. Ivanova

Creation of a Digital Model Twin of a Coal Mining Enterprise Based on Geospatial Data

Abstract. This article develops the concept of digital twins as applied to the coal mining industry. A comprehensive technological approach to the creation a Coal Mining Enterprise digital twin, combining aerial photography using an unmanned aerial vehicle, geodetic measurements, photogrammetric processing, and machine learning is proposed. The use of unmanned aerial vehicles is significantly more cost-effective, safer, and more efficient than ground-based methods for obtaining data on the surface of a quarry, open-pit mine, or outcrop mine. The high quality and accuracy of the resulting geospatial data enables the creation of up-to-date mining documentation, digital models, and digital twins, improved design efficiency, environmental impact assessments, and process forecasting. The results of integrating a digital twin into the "Drilling and Blasting Operations Planning and Monitoring System" web application are described. The proposed methodology is applicable to mine surveying and geodetic services, engineering surveys during quarry construction and operation, and the development of web applications and mining monitoring systems.

Keywords: coal mining, digital twin, aerial photography, unmanned aircraft, orthophotoplane, geodetic measurements, geospatial materials, 3D point model.

Аннотация. В статье разрабатывается концепция цифровых двойников применительно к угледобывающей промышленности. Предложен комплексный технологический подход к формированию цифрового двойника угледобывающего предприятия, объединяющий аэрофотосъемку с применением беспилотного воздушного судна, геодезические измерения, фотограмметрическую обработку и машинное обучение. Применение беспилотного воздушного судна значительно экономичнее, безопаснее и эффективнее наземных методов получения данных о поверхности карьера, разреза или рудника. Высокое качество и точность получаемых геопространственных данных позволяет создавать актуальную горно-графическую документацию, цифровые модели и двойники, повышать эффективность проектирования работ, проводить оценку воздействия на окружающую среду, прогнозировать течение различных технологических процессов. Описаны результаты интеграции цифрового двойника в web-приложение «Система планирования и мониторинга буровзрывных работ». Реализация предлагаемой методики применима для маркшейдерских и геодезических служб предприятий, инженерных изысканий при строительстве и эксплуатации карьеров, разработки web-приложений и систем мониторинга горных работ.

Ключевые слова: угледобывающая отрасль, цифровой двойник, аэрофотосъемка, беспилотное воздушное судно, ортофотоплан, геодезические измерения, геопространственные материалы, точечная 3D модель.

ВВЕДЕНИЕ

Современная отечественная угледобывающая промышленность находится в преддверии струк-

турной трансформации, которая может вывести ее на новый виток технологического развития. Переход от традиционной добычи к высокотехнологичной переработке и устойчивому производству является важной частью энергетической и промышленной

¹ Исследование выполнено в рамках мероприятия «Геоинформационная система цифрового регионального управления» комплексной научно-технической программы полного инновационного цикла «Разработка и внедрение комплекса технологий в областях разведки и добычи твердых полезных ископаемых, обеспечения промышленной безопасности, биоремедиации, создания новых продуктов глубокой переработки из угольного сырья при последовательном снижении экологической нагрузки на окружающую среду и рисков для жизни населения» (утвержденной распоряжением Правительства Российской Федерации от 11 мая 2022 г. № 1144-р).

безопасности страны. С этой целью формируются научно-технические центры, взаимодействующие с вузами и научными институтами, получают приоритетную поддержку научные проекты, результаты которых активно внедряются в работу угольных предприятий [1].

С помощью алгоритмов машинного обучения автономные буровые установки способны работать без вмешательства человека, а интеллектуальные комбайны – не только анализировать структуру угольных пластов, но и автоматически регулировать параметры разработки, что позволяет в значительной степени сократить потери угля, снизить затраты на техническое обслуживание и уменьшить влияние человеческого фактора [2].

Благодаря цифровым технологиям стало возможным обеспечивать безопасные условия труда, предотвращать аварии. Современные датчики способны измерять уровень вредных веществ, температуры, влажности и других параметров, которые влияют на безопасность. Системы прогнозной аналитики анализируют вибрации и структурные изменения горных пород и предсказывают возможные обрушения. Автоматизированные вентиляционные системы по данным сенсоров регулируют подачу воздуха, минимизируя скопление газов [2].

Беспилотные самосвалы и электровозы используются для транспортировки угля без участия человека, дроны – для мониторинга процесса выработок и обнаружения возможных угроз в виде трещин или скоплений газа. Это способствует повышению безопасности, а также сокращению затрат на эксплуатацию и техническое обслуживание техники [2].

Современные угольные предприятия широко используют технологии искусственного интеллекта для оптимизации различных процессов – от аналитики износа и простого оборудования до создания автоматизированной системы управления цепочками поставок.

Цифровые платформы обеспечивают оперативное распределение задач, проверку деятельности сотрудников, контроль за подачей и транспортировку угля в режиме реального времени, снижая риски для персонала [3].

Одним из новейших решений является создание цифровых аналогов (двойников) шахт. Это виртуальные модели, с помощью которых можно анализировать и регулировать производственные процессы [4]. Технология цифрового двойника предполагает создание виртуальной копии физического объекта, процесса или системы, которая достоверно отражает их основные характеристики, функционирование и взаимодействие с окружающей средой [5].

Технологии цифровых двойников используются в различных отраслях для проектирования, производства и эксплуатации различных деталей и конструкций, для проведения испытаний и моделирования высокотехнологичных процессов, для прогнозирования развития различных сценариев и проверки гипотез [5], активно применяются в машиностроении, энергетике, логистике, строительстве, в городских системах жизнеобеспечения [6].

В научных публикациях представлены различные определения понятия «цифровой двойник». Например, согласно работе Старка и Дамерау, цифровой двойник – это цифровой аналог (модель) определенного продукта или определенной системы продуктов и услуг, который воспроизводит его основные характеристики, свойства, состояния и поведение через информацию и данные, представленные в пределах одной или нескольких фаз жизненного цикла этого продукта [7]. Таким образом, цифровой двойник можно рассматривать, как виртуальную модель физического объекта, которая точно отражает его характеристики, поведение и взаимодействие с окружающей средой, помогает отслеживать состояние объекта, обнаруживать возможные сбои и принимать взвешенные решения относительно его обслуживания и эксплуатации [8,9].

В зависимости от сложности алгоритмов и функционала [10] цифровые двойники представлены достаточно простыми аналогами реальных физических объектов, содержащими математический аппарат, с помощью которого можно определить необходимые характеристики. Более продвинутые цифровые двойники при помощи алгоритмов машинного обучения могут получать, анализировать и передавать информацию, устанавливая отклонения от нормы, систематически обновляться при получении новых данных и, таким образом, способствовать усовершенствованию физического прототипа. Самые последние разработки цифровых двойников оснащены искусственным интеллектом, обучены моделировать разнообразные сценарии и прогнозировать поведение своего физического двойника [4,11].

Целью выполнения работ в рамках настоящего исследования являлось создание цифровой модели объекта угледобывающего предприятия как части web-приложения «Система планирования и мониторинга буровзрывных работ» на основе данных, полученных в ходе проведения аэрофотосъемки.

Научная новизна проекта заключается в использовании технологий искусственного интеллекта и машинного обучения в областях разведки и добычи твердых полезных ископаемых. Его реализация

направлена на создание сквозных инновационных решений для угольной отрасли – от разведки месторождений до переработки сырья. Практическая значимость определяется повышением эффективности угледобычи, а также снижением негативного воздействия на окружающую среду и рисков для жизни населения.

МЕТОДИКА ПОДГОТОВКИ ГЕОПРОСТРАНСТВЕННЫХ МАТЕРИАЛОВ

В качестве объекта работ определена территория угольного разреза Кемеровской области площадью

783,5 га (объект угольной промышленности). Для получения геопространственных материалов над территорией горного отвода разреза проводилась аэрофотосъемка (АФС) в октябре 2024 года. Было выполнено 12 полетов, общее количество фотоснимков составил 674 штук, общий объем фотоснимков – 6,9 ГБ. До и после проведения полетов были получены все необходимые разрешения и выполнены все необходимые требования.

Общий план территории объекта работ представлен на рисунке 1.

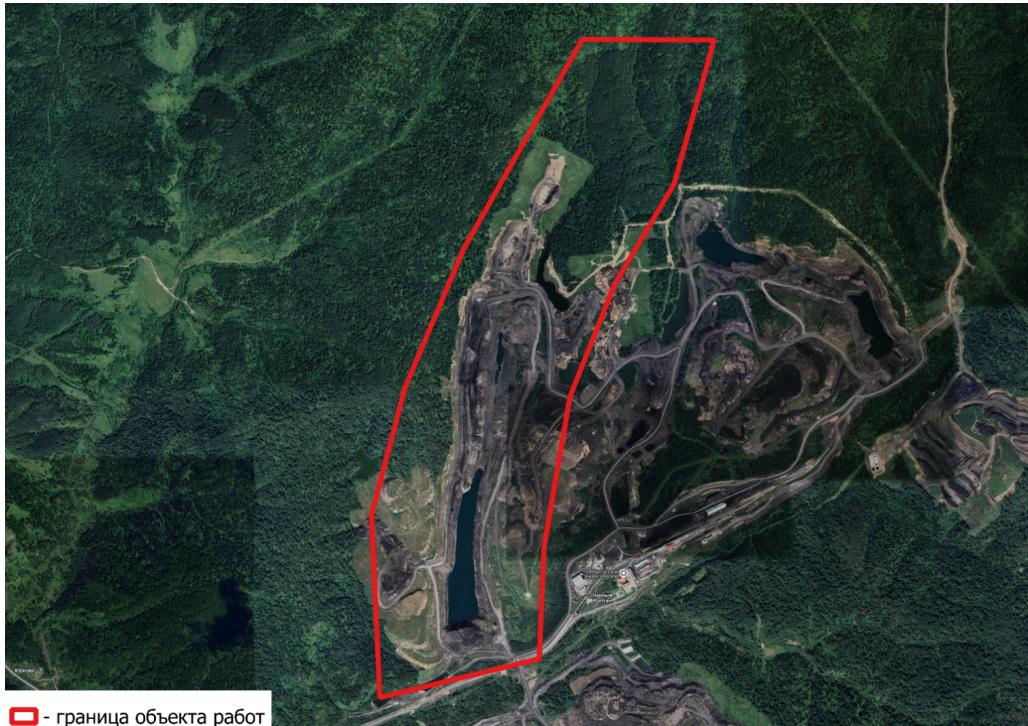


Рис. 1. Общий план территории объекта работ (фото из коллекции КемГУ)

Геодезические измерения

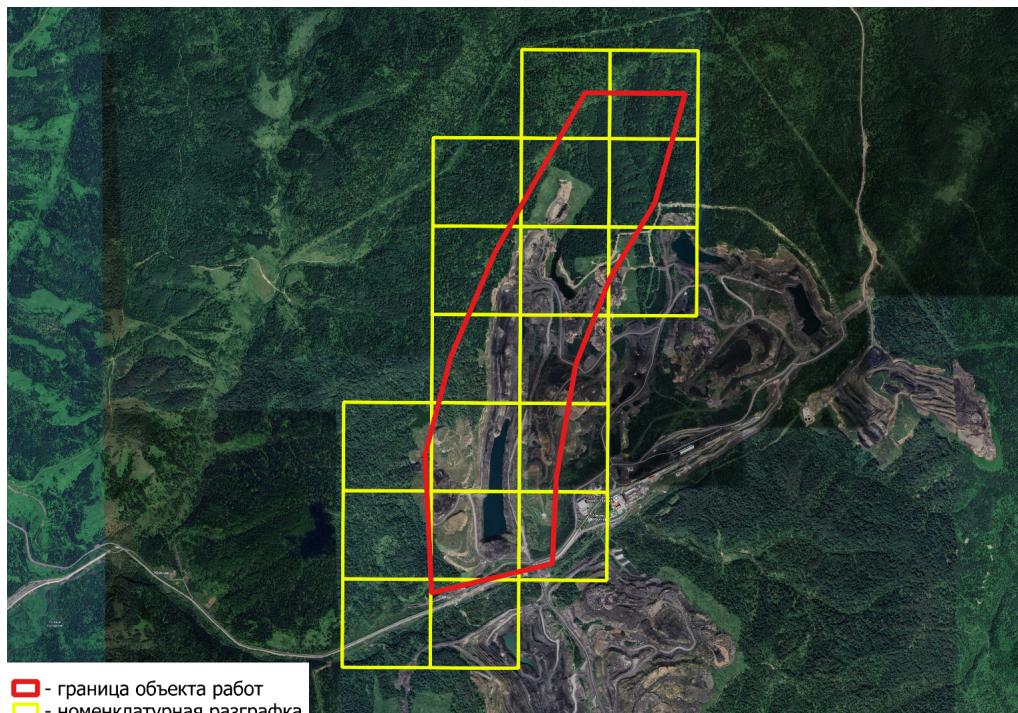
Геодезические измерения выполнялись с применением поверенных спутниковых геодезических приемников PrinCe i90, использованных в качестве роверов, и спутникового геодезического приемника PrinCe P5U, размещенного на точке базовой станции «КЕМЕ». Все использованные в работе геодезические приемники прошли метрологическую поверку, подтверждающую их годность к эксплуатации.

Расчет требуемого количества контрольных планово-высотных опознавателей для карт (планов) масштабом 1:2000 производился в соответствии с [12]: каждый отдельно уравниваемый блок был обеспечен контрольными опознавателями в количестве N_{KO} не меньшем, чем количество контрольных опознавателей, вычисленное по формуле:

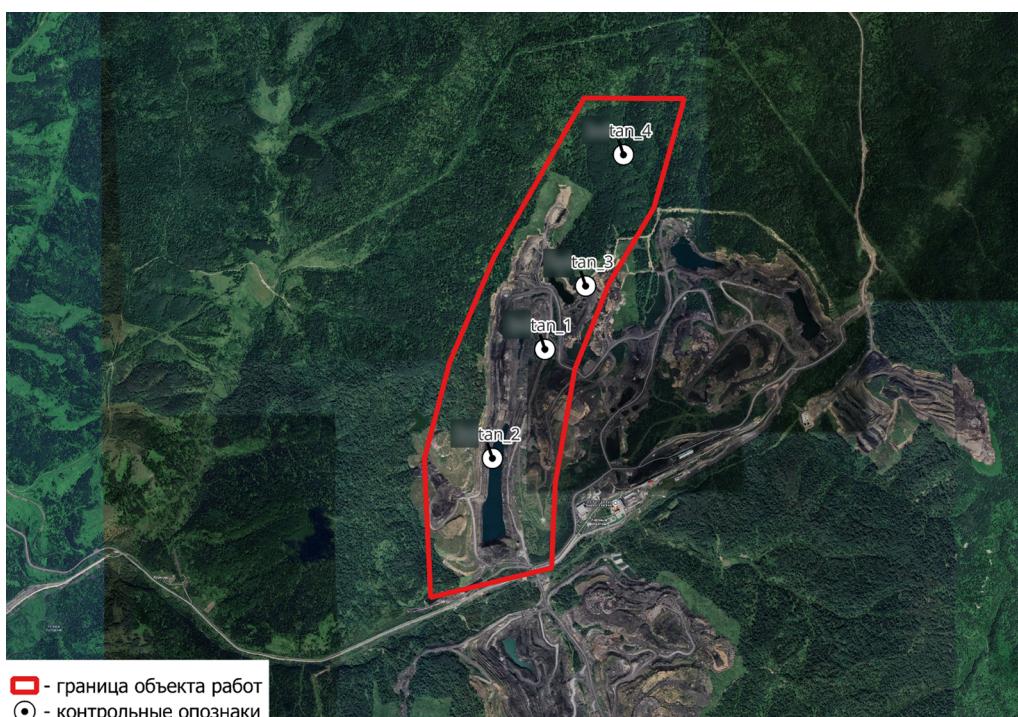
$$N_{KO} = \text{ОКРУГЛ}(N_{NL}/20+2,5), \quad (1)$$

где N_{NL} – число номенклатурных листов, покрываемых блоком.

Число номенклатурных листов определяли путем нанесения полигона, подлежащего съемке, на номенклатурную разграфку плана требуемого масштаба – 1:2000. Площадь одного номенклатурного листа составляла 25 га. Съемочный полигон накладывали на номенклатурную разграфку неравномерно (рисунок 2а), количество номенклатурных листов, взятых для расчета контрольных опознавателей, составило 18 шт. В работе было использовано 4 контрольных опознавателя (рисунок 2б), которые были равномерно распределены по площади съемочного блока.



(а)



(б)

Рис. 2. Номенклатурная разграфка (а) и расположение контрольных опознавательных знаков (б) на полигоне объекта работ (фото из коллекции КемГУ)

Обработка данных геодезических измерений

Обработка спутниковых геодезических измерений на пунктах планово высотного обоснования производилась в программном комплексе Trimble Business Center (Trimble Inc., Саннивейл, Калифорния, США). Определение координат и высот кон-

трольных опознавательных знаков производили относительно точки базовой станции.

Контроль точности обработанных данных геодезических измерений

Контроль точности обработанных данных геодезических измерений производился по среднеква-

дратической погрешности (СКП) определения планового и высотного положений контрольных опознавателей (таблица 1).

Таблица 1

**СКП определения координат и высот
контрольных опознавателей**

| Наименование пункта | СКП (план/высота), м |
|---------------------|----------------------|
| tan_1 | 0,015/0,020 |
| tan_2 | 0,012/0,016 |
| tan_3 | 0,008/0,013 |
| tan_4 | 0,020/0,027 |

По результатам спутникового определения координат контрольных опознавателей наибольшая СКП планового положения составила 0,02 м, что удовлетворяло нормативным требованиям [12].

Проведение аэрофотосъемки

Аэрофотосъемка выполнялась с применением беспилотного воздушного судна (БВС) мультироторного типа модели DJI Mavic 3 Enterprise, его технические характеристики приведены в таблице 2.

Оснащен БВС DJI Mavic 3 Enterprise был бортовым геодезическим приемником и штатной цифровой фотокамерой Hasselblad, технические характеристики которых приведены в таблице 3.

Таблица 2

Технические характеристики БВС DJI Mavic 3 Enterprise

| Характеристика | Значение |
|---|--|
| Тип БВС | мультироторный |
| Скорость полета (воздушная) | 5 м/с (режим С). 15 м/с (режим N). 21 м/с (режим S). |
| Продолжительность полета | до 45 минут |
| Максимальная высота полета над уровнем моря | до 6000 м |
| Диапазон рабочих температур | -10°C - +40°C |
| Максимальная взлетная масса | 1050 г |
| Размер по диагонали (без пропеллеров) | 380,1 мм |

Таблица 3

Технические характеристики полезной нагрузки DJI Mavic 3 Enterprise

| GNSS-приемник | |
|--|--|
| Характеристика | Значение |
| отслеживаемые сигналы | GPS, GLO, GAL, BEI: L1/L2 |
| точность определения положения (PPK) | По горизонтали: 1 см + 1 мм/км; По вертикали: 1.5 см + 1 мм/км. |
| Цифровая фотокамера Hasselblad | |
| Характеристика | Значение |
| максимальное разрешение снимков (m x n) | 20 Мегапикселей (5280 x 3956) |
| размер матрицы (a x b) | 18 x 13,5 мм |
| фокусное расстояние объектива фотокамеры (f) | 12 мм |
| тип затвора | глобальный |

Подготовка полетных заданий

Подготовка полетных заданий выполнялась в программном обеспечении TeoFly [12]. Проектное номинальное пространственное разрешение пикселя на местности (м/пиксель) рассчитали по формуле:

$$r_{pix} = \max \left(\frac{h_0 \cdot a}{f \cdot m} \vee \frac{h_0 \cdot b}{f \cdot n} \right), \quad (2)$$

где h_0 – проектная высота съемки над рельефом, м;
 a – размер матрицы (длина), мм;
 b – размер матрицы (ширина), мм;
 f – фокусное расстояние объектива, мм;
 m – максимальное горизонтальное разрешение снимков, пикселей;
 n – максимальное вертикальное разрешение снимков, пикселей.

r_{pix} составило 7,2 см при высоте съемки над рельефом 253 м. Продольное и поперечное перекрытия составили 79% и 60% соответственно.

Камеральная обработка данных аэросъемок

Фотограмметрическая обработка

Работы по созданию цифровых геопространственных материалов по данным АФС выполнялись в программе Agisoft Metashape Professional (ООО «ЖИВОЙСОФТ», Москва, Россия).

Исходными материалами для создания этих продуктов послужили оригинальные, необработанные цифровые аэрофотоснимки. Использовались фокусное расстояние, настройки модели камеры (элементы внутренней ориентации фотографий в Agisoft Metashape Professional), а также координаты и высоты центров проекции (элементы внешней ориентации) и контрольных точек (контроль точности) аэрофотоснимков.

Были выполнены ортотрансформирование аэрофотоснимков в соответствии с параметрами внешней ориентации и автоматическая идентификация изображения, построена цифровая модель местности с разреженными точками. Для местности была создана модель с разреженными и плотными точками (3D-точечная модель) с контролем положения, а также 2D-карта рельефа и цифровая ортофотоплоскость с корректировкой линии разреза для улучшения интерпретации характерных особенностей местности. Для создания цифровых геопространственных продуктов была проведена фотограмметрическая обработка.

Контроль фотограмметрической обработки

В профессиональном программном обеспечении Agisoft Metashape отслеживался процесс выравнивания аэрофотоснимков на основе получен-

ных точек проекции. Местоположение планируемых высотных контрольных точек, полученных в результате фотограмметрической обработки, было определено путем вычисления разницы в координатах и высотных отметках на ортоправленных спутниковых геодезически обработанных изображениях. Величина ошибок рассчитанных позиций контрольных точек (средних значений погрешностей) по результатам контроля составляла в плановом положении 0,038 м (Ошибка XY), в высотном – 0,078 м (Ошибка Z), что полностью удовлетворяет требованиям, указанным выше [12].

Модель цифрового двойника, который становится инструментом планирования, анализа и контроля производственных процессов, представляет собой виртуальную копию участка месторождения, полученную по итогам обработки материалов **БВС** [13]. В последовательный бизнес-процесс по получению плана горных работ посредством цифрового двойника угольного месторождения (рисунок 3) входит:

- **сбор данных с БВС** (в результате формируется массив снимков с высоким пространственным разрешением, обеспечивающий основу для дальнейшей фотограмметрической обработки, продолжительность этапа – до 6 часов);

- **автоматизированная фотограмметрическая обработка** (полученные изображения подвергаются камеральной обработке с использованием специализированного программного обеспечения, создаётся ортофотоплан участка и его облако точек высокой плотности до 200 млн элементов, с пространственным шагом около 2 см; продолжительность – 6–8 часов.);

- **итоговая обработка данных** (целью этапа является оптимизация объёма данных для дальнейшей загрузки в ПО инженера, выполняется прореживание облака точек до 0,5–1 млн элементов с варьируемым расстоянием между точками от 2 до 200 см, время выполнения – 1–2 часа);

- **добавление тематических слоёв** (осуществляется детализация модели, интегрируются слои, отражающие инфраструктурные объекты, слои с геологической информацией, ранее полученные данные маркшейдерской съемки, продолжительность – 6–8 часов);

- **передача результатов заказчику** (итоговый цифровой двойник размещается в ПО инженера угольного предприятия);

- **использование цифрового двойника.**

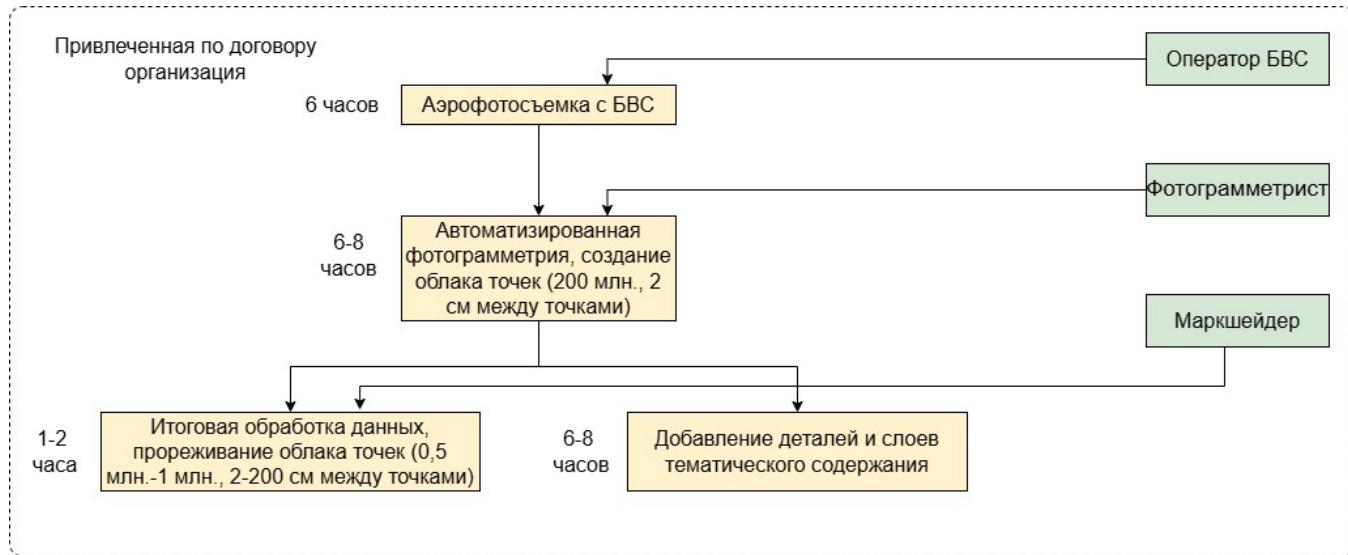


Рис. 3. Бизнес-процесс по получению плана горных работ угольного предприятия

Цифровой двойник размещён в web-приложении «Система планирования и мониторинга буровзрывных работ». В web-приложении с помощью цифрового двойника пользователь может выполнять следующие задачи: проектирование буровых работ, выполнение необходимых измерений, формирование трехмерных детализированных схем откосов, определение координат точек и геометрических размеров выемочного блока, построение профиля высот, формирование паспорта буровзрывных работ.

ФОРМИРОВАНИЕ ЦИФРОВОГО ДВОЙНИКА

Формирование цифрового двойника разреза основывается на интеграции разнородных геопространственных данных.

Так, цифровой двойник построен с использованием цифровых геопространственных данных, полученных с БВС: цифровой ортофотоплан, цифровая модель местности, облако точек. Исходными материалами для построения геопространственных данных служат: цифровые аэрофотоснимки, элементы внешнего ориентирования (точные центры фотографирования, опознавания), элементы внутреннего ориентирования аэрофотокамеры (фокусное расстояние, значение дисторсий и др.), контрольные опознавания, траектория движения воздушного судна, данные инерциально-измерительной системы, измерения лазерного лидара.

В контексте цифрового двойника разреза ортофотоплан используется как текстурная картографическая подложка, обеспечивающая визуализацию

актуального состояния объекта: уступов, инфраструктуры, растительности. Кроме того, ортофотопланы различных временных интервалов позволяют осуществлять мониторинг динамики горных работ. Цифровая модель местности (ЦММ) отражает трёхмерную геометрию рельефа. Для участка месторождения ЦММ является ключевым инструментом геометрического анализа: она используется для расчёта объёмов выемки/насыпи, оценки устойчивости откосов, проектирования горнодобывающих работ.

Облако точек представляет собой набор пространственно привязанных трёхмерных координат, характеризующих объект с максимальной детализацией. В составе цифрового двойника облако точек выполняет функции высокоточного источника данных для построения 3D-моделей уступов, проведения измерений (углы откосов, протяжённость, объём выемки, определение координат). Высокая точность облака точек обеспечивает возможность анализа неточностей проектирования и оценки потенциальных рисков проведения буровзрывных работ.

На основании данных, полученных с помощью аэрофотосъемки, были сформированы цифровые геопространственные материалы – ортофотоплан объекта с пространственным разрешением 5,82 см/пикс, ЦММ (2D карта высот) пространственным разрешением 23,3 см/пикс, ЦММ в виде облака точек (точечная 3D модель) со средней плотностью 26,5 т/м².

Общий вид материалов представлен на рисунке 4.

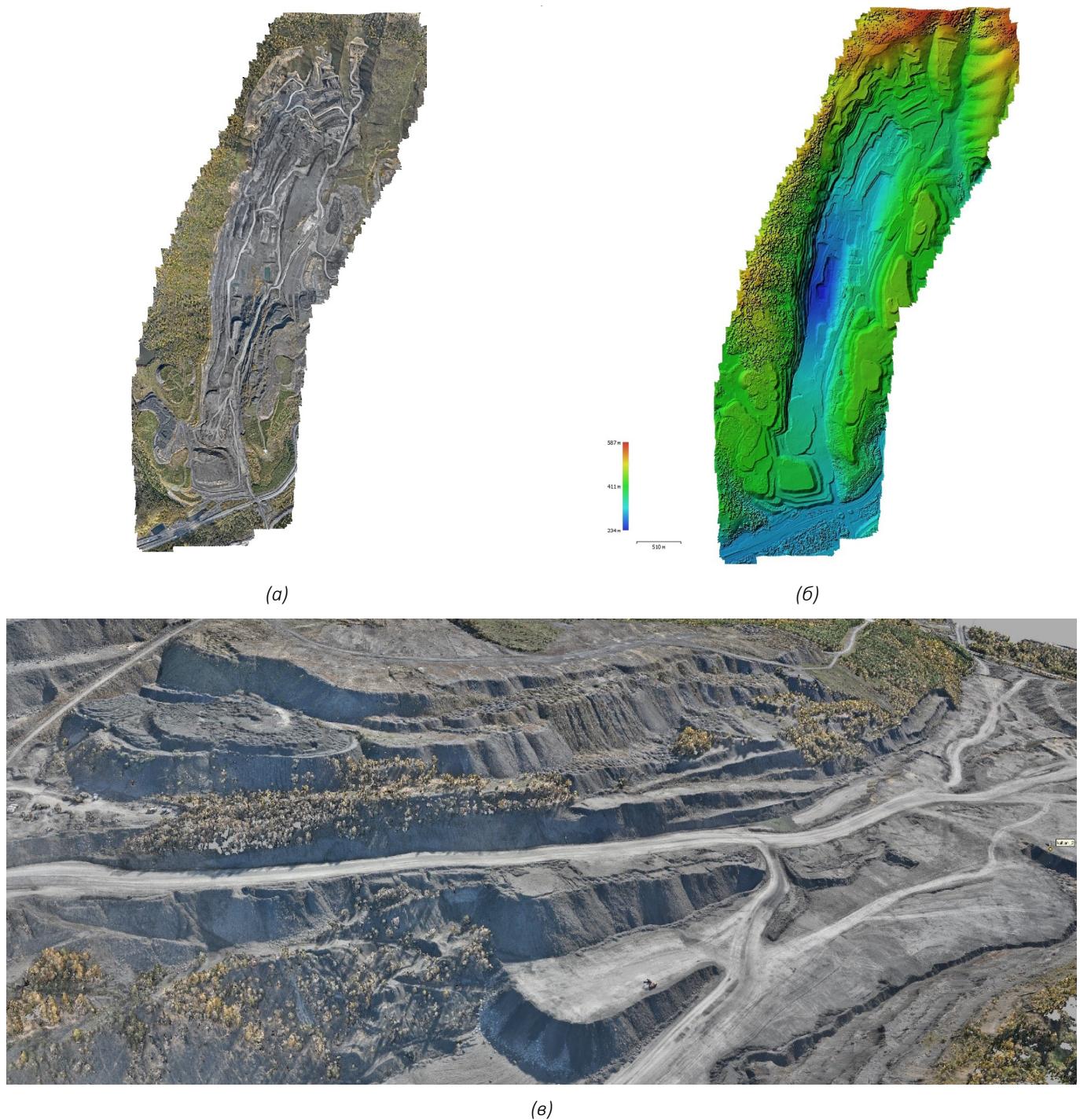


Рис. 4. Общий вид (а) цифрового ортофотоплана, (б) ЦММ (2D карта высот) и (в) ЦММ в виде облака точек (точечная 3D-модель) по данным АФС

СФЕРА ПРИМЕНЕНИЯ ЦИФРОВЫХ ДВОЙНИКОВ: ОБСУЖДЕНИЕ

Применение цифровых двойников в различных сферах экономики и производства помогает оптимизировать комплексные трудоемкие процессы, сокращать время на разработку и создание новых продуктов, повышать качество продуктов и услуг, снижать затраты.

В строительной отрасли с помощью цифровых двойников конструируют здания, объекты инфраструктуры, что в дальнейшем обеспечивает их более эффективную эксплуатацию, а также сниженные риски и затраты при строительстве [14,15].

В транспортном секторе и в управлении грузоперевозками с помощью цифровых моделей можно управлять транспортными средствами, рассчитывать оптимальные маршруты грузоперевозок, осу-

ществлять мониторинг состояния грузов в режиме онлайн. Это позволяет снижать топливные затраты, сокращать время доставки груза, повышать качество обслуживания клиентов [16].

Применение цифровых двойников в работе предприятий энергетической промышленности позволяет моделировать возможные сценарии, оптимизировать потоки мощности, ликвидировать аварийные ситуации, обеспечивать надежность и эффективность работы энергосистем [17].

В угледобывающей отрасли модели цифровых двойников применяются для мониторинга и управления оборудованием, оптимизации транспортировки угля, планирования и прогнозирования спроса, повышения эффективности поставок, моделирования и оптимизации производственных процессов, использования ресурсов, управления безопасностью, снижения негативного воздействия на окружающую среду [18,19].

Компания Кемеровской области создала цифровой двойник Талдинского угольного разреза по данным геологоразведочных работ. Были созданы цифровые модели скважин, породных отвалов и участков месторождений, высокоперспективных для освоения. Это позволило более тщательно планировать весь комплекс работ, оптимизировать учет запасов и их движения, усовершенствовать геолого-маркшейдерское сопровождение [20].

Цифровые двойники, получаемые с помощью аэрофотосъемки, намного подробнее и точнее моделей, построенных по наземным съемкам.

Аэрофотосъемка с использованием БВС мультироторного типа, благодаря своей маневренности и доступности, стали популярным инструментом для различных областей применения.

Преимущества аэрофотосъемки с БВС мультироторного типа:

1. Доступность: более низкая стоимость амортизации за счет большего ресурса и меньшей стоимости.

2. Мобильность: легкость транспортировки, быстрая подготовка к работе.

3. Высокое качество изображений: возможность получения более качественных аэрофотоснимков за счет стабилизации БВС и внесения изменений в настройки камеры в полете.

4. Гибкость: возможность съемки в труднодоступных местах, где взлет БВС самолетного типа невозможен.

В результате аэрофотосъемки и лазерного сканирования с БВС получены сверхточные цифровые

геопространственные материалы, превосходящие по качеству данные, получаемые традиционными методами. Это способствовало оптимизации дальнейших работ на одном из предприятий угольной промышленности Кемеровской области – Кузбасса [13].

С помощью 3D-модели, созданной по материалам аэрофотосъемки и лазерного сканирования с БВС, была проведена оценка параметров раз渲ала взрывной горной массы. Данные, полученные на одном из угольных разрезов Кузбасса, позволяют точно измерить ориентацию и длину трещин на обширной территории, зачастую опасной либо вообще недоступной для ручного картирования [13].

Традиционный метод наземной съемки блока с помощью GNSS-приемника подразумевает получение около 125 точек с шагом 20 м. При таком методе получения геоданных отклонение фактической высотной отметки устья скважины может составлять более двух метров. Трехмерная модель рельефа, построенная по данным съемки с БВС, содержит закоординированное плотное облако точек (до 3 млн точек) с шагом 20 см. Такая точность снижает риски ошибок при проектировании и контроле результатов проводимых работ [21].

ЗАКЛЮЧЕНИЕ

В ходе выполнения работ было подготовлено планово-высотное обоснование и осуществлен контроль его точности, произведена аэрофотосъемка объекта работ с использованием БВС DJI Mavic 3 Enterprise, выполнена разработка и проверка точности цифровых геопространственных материалов.

По результатам обработки данных были получены цифровые геопространственные материалы: цифровой ортофотоплан с пространственным разрешением 5,82 см/пикс, ЦММ объекта работ в виде плотного облака точек (точечная 3D модель) со средней плотностью 26,5 т/м².

Совокупное использование ортофотоплана, цифровой модели местности и облака точек позволяет сформировать комплексный цифровой двойник участка угледобычи. Интеграция этих данных обеспечивает решение широкого спектра задач – от оперативного мониторинга до стратегического планирования и оптимизации горнодобывающих процессов.

СПИСОК ЛИТЕРАТУРЫ

1. Черное золото: почему российская угольная отрасль способна стать одним из локомотивов промышленного роста [Электронный ресурс]. URL: <https://www.gazeta.ru/> (Дата обращения: 30.07.2025).
2. Автоматизация угледобычи: как цифровые технологии меняют подход к шахтам [Электронный ресурс]. URL: <https://rosstip.ru/> (Дата обращения: 30.07.2025).
3. Yu R., Yang X., Cheng K. Deep learning and IoT enabled digital twin framework for monitoring open-pit coal mines // *Frontiers in Energy Research*. 2023. №11. 1265111. <https://doi.org/10.3389/fenrg.2023.1265111>.
4. Сосфенов Д.А. Цифровой двойник: история возникновения и перспективы развития // Интеллект. Инновации. Инвестиции. 2023. №4. С. 35–43.
5. Идеальная копия: какие задачи решают цифровые двойники в разных отраслях [Электронный ресурс]. URL: <https://sber.pro/> (Дата обращения: 31.07.2025).
6. Денисов С.Г. Расширение сферы применения цифровых двойников // БИТ. 2024. №4 (32). С. 18-21.
7. Stark R., Damerau T. Digital Twin // CIRP Encyclopedia of Production Engineering. 2019. № 66. Р. 1–8. https://doi.org/10.1007/978-3-642-35950-7_16870-1.
8. Абрамов В.И., Гордеев В.В., Столяров А.Д. Цифровые двойники: характеристики, типология, практики развития // Вопросы инновационной экономики. 2024. Т. 14, № 3. С. 691 – 716.
9. Liu M., Fang S., Dong H., Xu C. Review of digital twin about concepts, technologies, and industrial applications // *Journal of Manufacturing Systems*. 2020. № 58. Р. 346–361.
10. Картер Д. Синтез данных и цифровые двойники. 2024. 185 С.
11. Мальцева А. Что такое digital twin и как они помогают бизнесу экономить миллионы [Электронный ресурс]. URL: <https://www.raiffeisen-media.ru/> (Дата обращения: 01.08.2025).
12. ГОСТ Р 59562-2021. Съемка аэрофототопографическая. Технические требования.
13. Петерс К.И., Щербакова Л.Н., Федулова Е.А., Кузнецов А.Д., Бурмин Л.Н. Практика применения 3D-модели в горной промышленности на примере угольных предприятий в Кемеровской области – Кузбассе // Уголь. 2024. №10. С. 110–117.
14. Голядкина А.Д., Субботин Д.П., Красникова А.Н., Бугаевский Д.О. Цифровые двойники в строительстве // Градостроительство. Инфраструктура. Коммуникации». 2023. №4(33). С. 17–21.
15. Jiang Y., Li M., Guo D., Wu W., Zhong R.Y., Huang G.Q. Digital twin-enabled smart modular integrated construction system for on-site assembly // *Computers in Industry*. 2022. №136. 103594. <https://doi.org/10.1061/9780784485262.055>.
16. Абрамов В.И., Абрамов О.В., Поливанов К.В., Семенков К.Ю. Особенности создания цифровых двойников управления логистикой предприятий // Новое в экономической кибернетике. 2024. № 2. С. 52-61.
17. Heluany J.B., Gkioulos V. A review on digital twins for power generation and distribution // *International Journal of Information Security*. 2023. № 2. Р. 1171–1195.
18. Wang Y., Wang L., Wang H., Li R., Li W. Data-Driven and Model-Driven Integration Approach for Optimizing Equipment Safety Investment in Digital Twin Coal Mining Enterprises // *Applied Sciences*. 2024. №14(23). 11101. <https://doi.org/10.3390/app142311101>.
19. Жданеев О.В., Власова И.М. Вызовы и приоритеты цифровой трансформации угольной отрасли // Уголь. 2023. № 1. С. 62–69.
20. Рада А.О., Петерс К.И., Кузнецов А.Д. Проектирование буровзрывных работ на основе 3D-модели выемочного блока // Уголь. 2024. №6. С. 87–91.

УДК: 621.31, 004.8

Особенности удаленного видеомониторинга энергетической инфраструктуры арктической зоны¹

D.V. Serdechnyj, N.S. Akinshin, V.D. Kutkov

Features of Remote Video Monitoring of Energy Infrastructure in the Arctic zone

Abstract. This paper provides an analytical review of the Arctic energy infrastructure, examining the types of facilities most susceptible to wear and external impacts and systematizing modern monitoring approaches. The architectures of convolutional neural networks (CNNs) used for defect detection, segmentation, and classification are analyzed. Requirements for constructing training datasets are defined, including the specifics of dataset formation under conditions of low illumination, icing, and limited scene contrast. The principles of integrating video monitoring systems into energy complexes using autonomous computing modules (edge devices) and thermal imaging sensors are described. The discussed methods and technical solutions form a basis for developing intelligent control and diagnostic systems for energy facilities in the Arctic, thereby improving the reliability of infrastructure condition assessment, reducing the probability of accidents, and optimizing operational processes in regions with limited access to traditional technical supervision tools.

Keywords: Arctic, energy infrastructure, video monitoring, neural networks, convolutional neural networks (CNNs), diagnostics, artificial intelligence (AI), sustainable development, detection, computer vision.

использованием автономных вычислительных модулей и тепловизионных сенсоров. Рассмотренные методы и технические решения создают основу для формирования интеллектуальных систем контроля и диагностики энергетических объектов в Арктике, обеспечивая повышение достоверности оценки состояния инфраструктуры, снижение вероятности аварийных ситуаций и оптимизацию эксплуатационных процессов в регионах с ограниченным доступом к традиционным средствам технического надзора.

Ключевые слова: Арктика, энергетическая инфраструктура, видеомониторинг, нейросети, сверточные сети, диагностика, искусственный интеллект, устойчивое развитие, детектирование, компьютерное зрение.

ВВЕДЕНИЕ

Энергетическая инфраструктура арктических территорий Российской Федерации представляет собой совокупность объектов электро- и теплоэнергетики, обеспечивающих функционирование промышленных комплексов, транспортных узлов и населенных пунктов в условиях ограниченной доступности и низких температур. Территориальная изолированность, высокая стоимость энергоресур-

сов и деградация инженерных сооружений, вызванная изменением геоклиматических условий, создают необходимость перехода к технологиям непрерывного технического контроля и дистанционной диагностики [1].

Большая часть объектов располагается вне зон централизованного электроснабжения и функционирует в режиме автономных энергоузлов. При этом эксплуатация линий электропередачи, подстанций, трубопроводов и тепловых сетей сопряжена с воздействием циклических температурных

¹ Статья подготовлена по результатам исследований, выполненных за счет средств федерального бюджета по государственному заданию «Научные, методологические и практические основы разработки и применения цифровых и интеллектуальных технологий в целях обеспечения устойчивого развития регионов Российской Федерации, включая удаленные и труднодоступные территории Сибири, Дальнего Востока и Арктической зоны»; код научной темы, присвоенной учредителем, – FZNW-2025-0021.

Д.В. Сердечный¹Н.С. Акиньшин²В.Д. Кутков³

¹Кандидат технических наук, начальник лаборатории цифровых и интеллектуальных технологий для развития территорий Российской Федерации, Государственный университет управления.

E-mail: dv_serdechnyj@guu.ru

²Аспирант, младший научный сотрудник лаборатории реверсивного инжиниринга, Государственный университет управления.

E-mail: ns_akinshin@guu.ru

³Аспирант, младший научный сотрудник лаборатории реверсивного инжиниринга, Государственный университет управления.

E-mail: vd_kutkov@guu.ru

деформаций, ветровых нагрузок, коррозии и механического износа [2]. Риски аварийных отключений и утечек возрастают в периоды полярной ночи, когда выполнение традиционных методов визуального контроля становится затруднительным.

Существующие формы мониторинга, основанные на инспекциях, тепловизионных обследованиях и периодических измерениях, не обеспечивают требуемой оперативности при обслуживании протяжённых сетей и распределённых энергообъектов [3]. Интенсификация процессов цифровизации в энергетике создала предпосылки для применения автоматизированных систем видеомониторинга, использующих методы анализа изображений и видеопотоков на основе нейросетевых моделей. Технологии сверточных нейронных сетей обеспечивают извлечение информативных признаков, необходимых для идентификации структурных нарушений, дефектов изоляции и механических повреждений оборудования. Интеграция таких моделей в автономные вычислительные модули обеспечивает непрерывный контроль оборудования и передачу диагностических данных в централизованные системы управления [4].

Цель исследования заключается в рассмотрении возможностей применения нейросетевых моделей видеомониторинга для автоматизированного контроля и диагностики объектов энергетической инфраструктуры арктических регионов России. Исследование включает анализ функционирования энергетических систем в условиях низких температур, систематизацию методов визуального контроля и выбор архитектурных решений, устойчивых к сложным климатическим и эксплуатационным условиям.

Методологическая основа исследования базируется на системном анализе и классификации технических решений в области автоматизированного мониторинга. Информационной базой служат научные публикации по тематике компьютерного зрения и диагностики, отраслевые отчёты по эксплуатации энергетических объектов в Арктической зоне РФ, а также нормативно-техническая документация. Исследование не предполагает проведения натурных экспериментов и фокусируется на аналитическом обобщении и систематизации подходов для формирования концептуальной модели применения нейросетевых технологий.

КАТЕГОРИИ ОБЪЕКТОВ ЭНЕРГЕТИЧЕСКОЙ ИНФРАСТРУКТУРЫ

В рамках работы выполняется декомпозиция энергетической инфраструктуры Арктики на классы

объектов, различающиеся по функциональному назначению, конструктивным особенностям и профилю эксплуатационных рисков. Выделяются следующие основные категории:

1. **Объекты генерации:** автономные дизельные электростанции (ДЭС), теплоэлектроцентрали (ТЭЦ), атомные станции, включая плавучие (ПАТЭС), и установки на основе возобновляемых источников энергии (ВИЭ).

2. **Объекты передачи и распределения:** воздушные и кабельные линии электропередачи (ЛЭП), трансформаторные подстанции и распределительные устройства [5].

3. **Объекты топливной инфраструктуры:** магистральные и локальные трубопроводы, объекты хранения топлива и регазификационные комплексы.

Для визуализации данной структуры предусмотрена разработка классификационной схемы, отражающей иерархию и взаимосвязи между элементами энергетической системы в арктическом контексте. Классификационная схема объектов энергетической инфраструктуры представлена на рисунке 1.

МЕТОДЫ ДИАГНОСТИКИ СОСТОЯНИЯ ОБЪЕКТОВ ЭНЕРГЕТИЧЕСКОЙ ИНФРАСТРУКТУРЫ

При выполнении диагностики традиционные подходы (визуальный осмотр, тепловизионный контроль, неразрушающий контроль) сочетаются с возможностями систем видеомониторинга. Для последних рассматриваются три основные задачи компьютерного зрения, применимые к диагностике технического состояния:

• **детектирование:** локализация дискретных дефектов на изображениях (например, повреждённые изоляторы, обрывы проводов, деформации опор);

• **сегментация:** пописельное выделение областей с аномалиями, не имеющих чётких геометрических границ (участки коррозии, зоны обледенения, утечки технологических жидкостей);

• **классификация:** определение состояния объекта или его компонента в целом (норма, предварийное состояние, авария).

Для адаптации к указанным задачам проводится анализ архитектур свёрточных нейронных сетей (CNN) и трансформеров зрения (ViT). Однoproходные детекторы семейства YOLO (You Only Look Once) [6] обеспечивают высокую производительность — в частности, YOLOv8 демонстрирует скорость обработки до 65 кадров в секунду при средней точности детектирования mAP = 0.82 на внутреннем тестовом наборе изображений линий электропередач [7].

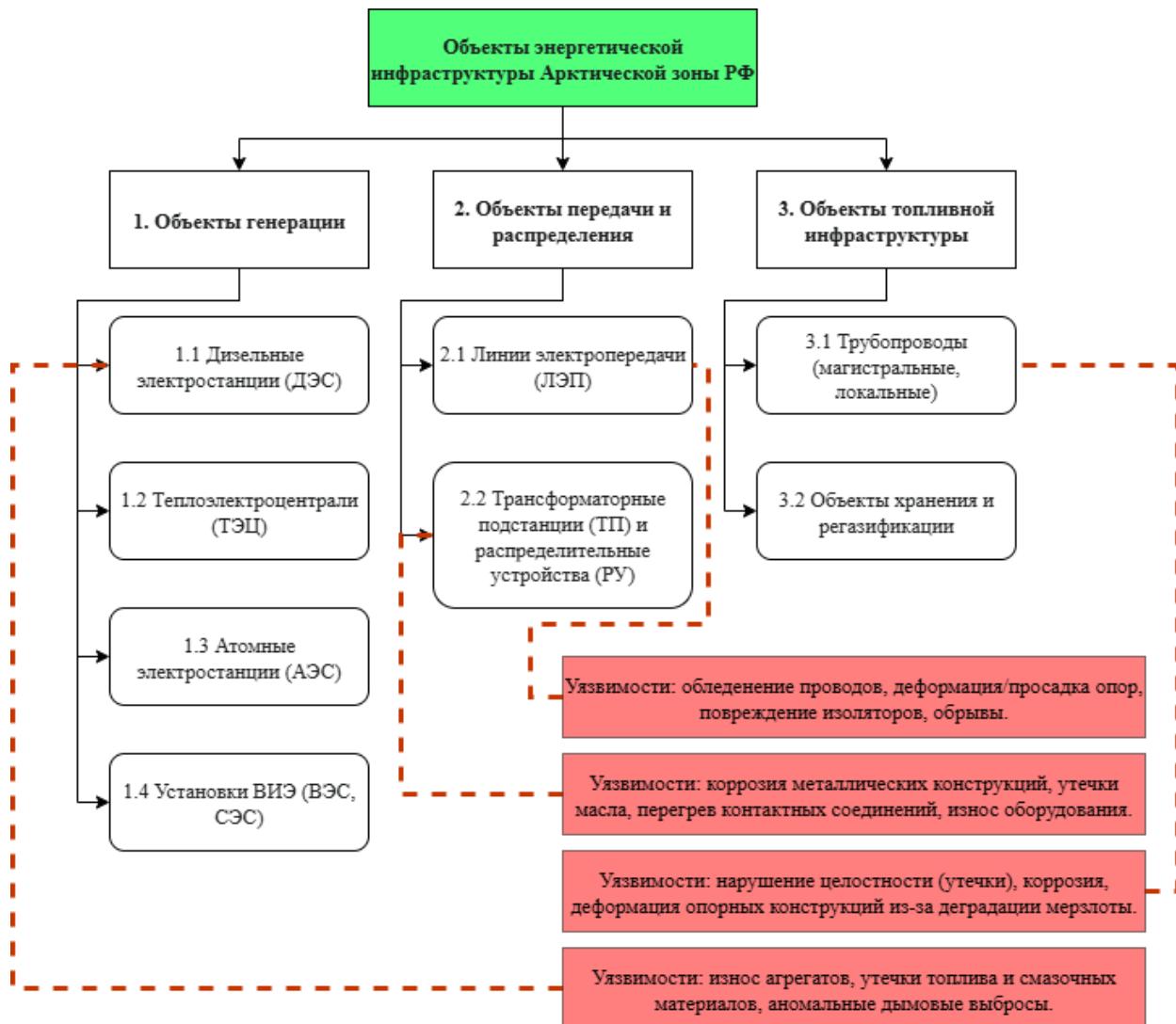


Рис. 1. Схема объектов энергетической инфраструктуры

Для формализации зависимости качества детектирования от характеристик данных и архитектуры модели применим обобщённое выражение, носящее качественный характер и описывающее общие тенденции, а не строгую функциональную зависимость:

$$Q = f(N_d, R_i, C_m, A_t), \quad (1)$$

где:

Q – интегральный показатель точности (например, mAP);

N_d – объём обучающей выборки;

R_i – разрешающая способность входных изображений;

C_m – вычислительная сложность модели (в количестве параметров, FLOPs);

A_t – уровень адаптации модели к внешним фактам (освещённость, шум, погодные искажения).

При увеличении N_d и A_t показатель точности Q растёт до насыщения, после чего прирост ми-

нимален, что подтверждает принцип «затухающих улучшений» при масштабировании данных.

Для задач, требующих более точной локализации дефектов, применяются двухходовые архитектуры (семейство R-CNN), которые достигают mAP до 0.90, однако при снижении скорости анализа до 10–15 fps [8]. Для задач сегментации применяются архитектуры типа U-Net и Mask R-CNN, обеспечивающие попиксельное выделение областей дефектов (например, коррозии, обледенения или участков износа).

В отличие от сегментации, детекция аномалий ориентирована на выявление нетипичных состояний оборудования без заранее заданных классов дефектов и используется при диагностике редких или неизвестных повреждений. Такие модели, как Autoencoder и GAN-based, определяют отклонения по статистическим признакам изображения, а не по заранее размеченным областям.

ИНСТРУМЕНТЫ ДИАГНОСТИКИ ОБЪЕКТОВ

Для систематизации выбора технических решений формируется таблица 1, связывающая типы

характерных для Арктики дефектов (механические повреждения, коррозия, обледенение, просадка опор вследствие деградации мерзлоты) с задачами видеоанализа и релевантными нейросетевыми архитектурами.

Таблица 1

Применение нейросетевых моделей для диагностики объектов энергетической инфраструктуры в Арктике

| Объект / Система | Целевой дефект / аномалия | Задача видеоанализа | Нейросетевая архитектура НС | Ключевое преимущество | Источник данных / тип съёмки |
|------------------------------------|--|--------------------------------|-----------------------------|---|---|
| Линии электропередачи (ЛЭП) | Обледенение проводов и изоляторов | Сегментация | U-Net / DeepLabv3+ | Точность контуров, количественная оценка | Беспилотное воздушное судно (БВС) (RGB, высокое разрешение) |
| | Повреждение изолятора (скол, пробой) | Детектирование | YOLOv8 / EfficientDet | Возможность анализа в режиме реального времени (≥ 30 fps) | БВС (RGB, оптический зум) |
| Опоры / фундаменты | Просадка, деформация (из-за таяния мерзлоты) | Детекция изменений | YOLO + Keypoint Detection | Анализ динамики, высокая точность смещений | Стационарная камера / Геопривязанные снимки |
| Подстанции (оборудование) | Поверхностная коррозия | Сегментация | Mask R-CNN | Точная оценка площади поражения | Ручная съёмка (RGB, HDR) |
| | Утечка трансформаторного масла | Детекция аномалий | Autoencoder / U-Net | Выявление нетипичных состояний | Стационарная камера (RGB) |
| | Локальный перегрев контактных соединений | Мультимодальное детектирование | Fusion CNN (YOLO-based) | Низкий уровень ложных срабатываний | Стационарная камера (RGB+IR) |
| Трубопроводы / опорные конструкции | Нарушение целостности, утечка продукта | Детекция аномалий | YOLO / Autoencoder | Быстрое сканирование протяжённых участков | БВС (RGB/IR), спутниковые снимки |
| Объекты генерации (ТЭЦ, ДЭС) | Аномальные выбросы (цвет, плотность дыма) | Классификация состояния | EfficientNet / ResNet | Интегральная оценка работы агрегата | Стационарная камера (RGB) |
| Установки ВИЭ (Ветрогенераторы) | Обледенение лопастей, трещины | Сегментация / детектирование | U-Net / YOLO | Бесконтактная диагностика труднодоступных узлов | БВС / Стационарная камера (RGB) |

Особое внимание в методологии уделяется учёту специфических условий сбора визуальных данных в Арктике. Анализируются факторы, влияющие на качество изображений: низкая и переменчивая освещённость (полярный день/ночь), наличие атмосферных осадков (снег, туман), блики от снежного и ледового покрова, ограниченный контраст сцены. В качестве методологического инструмента для адап-

тации моделей к данным условиям рассматриваются техники аугментации обучающих выборок, включая фотометрические преобразования, добавление синтетического шума и артефактов, имитирующих атмосферные помехи.

Итоговая методика сводится к последовательности аналитических процедур, которая представлена в виде обобщённой схемы на рисунке 2.

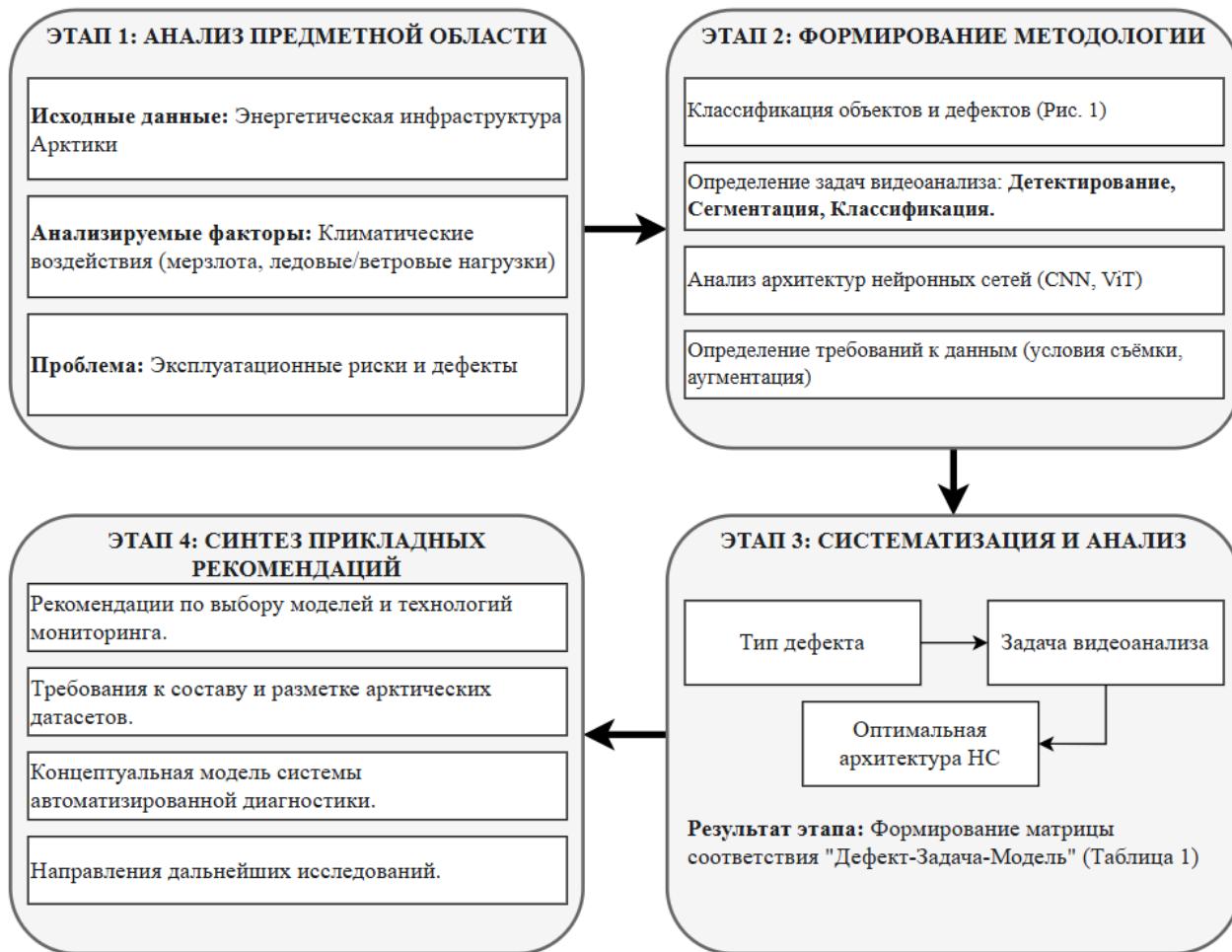


Рис. 2. Этапы исследования

Данная схема отображает логику перехода от анализа предметной области (инфраструктура и её уязвимости) к синтезу прикладных рекомендаций по выбору и адаптации нейросетевых моделей видеомониторинга.

Основой нейросетевого анализа изображений является применение свёрточных нейронных сетей (CNN), реализующих принцип иерархического извлечения признаков. Данный подход позволяет модели автоматически, в процессе обучения, формировать внутренние представления визуальных данных, переходя от низкоуровневых признаков (границы, градиенты, текстуры) к высокоуровневым (сложные

геометрические формы, части объектов). Такая иерархическая структура обеспечивает робастность к вариациям масштаба, ракурса и условий освещения, что является фундаментальным требованием для систем мониторинга, функционирующих в нестабильной внешней среде.

Для решения прикладных задач диагностики инфраструктуры используются специализированные архитектуры CNN. Для задач детектирования, целью которых является локализация дефектов с помощью ограничивающих рамок (bounding box), применяются однопроходные (single-shot) архитектуры, такие как YOLO, обеспечивающие высокую скорость об-

работки видеопотока. В задачах, требующих повышенной точности идентификации малоразмерных объектов, могут использоваться двухпроходные модели, например, семейства R-CNN [9]. Для задач сегментации, где необходимо попиксельное выделение областей аномалий (коррозия, обледенение), применяются энкодер-декодерные архитектуры, в частности U-Net, которая демонстрирует высокую эффективность при обучении на ограниченных наборах данных [10].

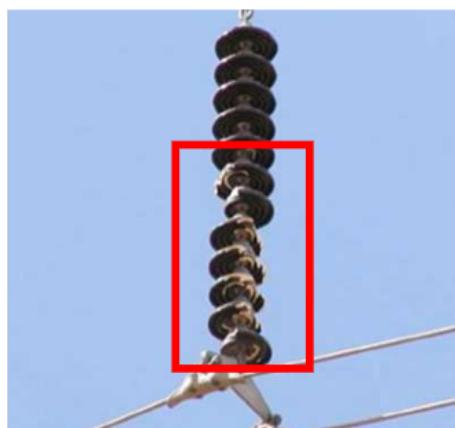
ФОРМИРОВАНИЕ СПЕЦИАЛИЗИРОВАННЫХ ДАТАСЕТОВ

Эффективность нейросетевых моделей напрямую определяется репрезентативностью и качеством обучающих выборок. Стандартные публичные датасеты не содержат целевых классов объектов и дефектов, характерных для энергетической инфраструктуры, и не отражают специфику арктических визуальных сцен. Следовательно, возникает необ-

ходимость в формировании специализированных датасетов. Процесс подготовки данных включает разметку, где для каждого дефекта указывается его локализация посредством ограничивающей рамки или попиксельной маски [11]. Примеры размеченных изображений дефектов энергетической инфраструктуры представлены на рисунке 3.

В условиях дефицита реальных размеченных данных, характерного для узкоспециализированных промышленных задач, ключевую роль играет процесс аугментации — искусственного расширения обучающей выборки [12]. Аугментация позволяет повысить вариативность данных и развить у модели инвариантность к определённым искажениям [13]. Для адаптации к арктическим условиям применяются следующие группы преобразований:

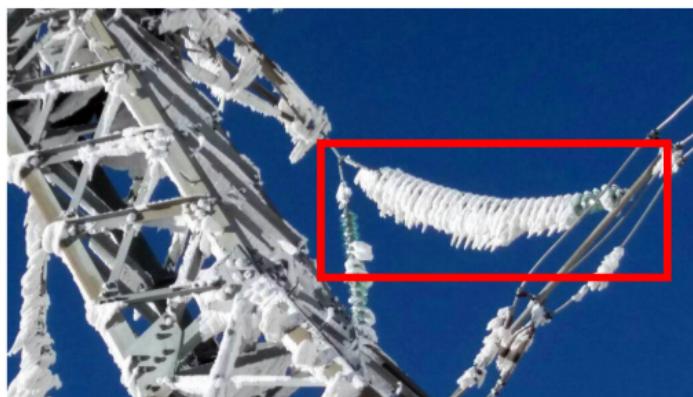
- **Фотометрические преобразования:** изменение яркости, контрастности, насыщенности для моделирования условий полярного дня и ночи, а также сумеречного освещения.
- **Имитация атмосферных явлений:** алгоритмическое добавление на изображения синтети-



а)



б)



в)



г)

Рис. 3. Примеры разметки дефектов на изображениях: а) детектирование повреждения изолятора; б) сегментация коррозии; в) сегментация обледенения; г) детектирование перегрева в ИК-спектре.

ческого снега, тумана, дождя и бликов от низкого солнца или ледового покрова.

• **Геометрические преобразования:** случайные повороты, масштабирование, сдвиги и аффинные искажения для моделирования различных ракурсов съёмки с БВС или стационарных камер.

Данный процесс позволяет на основе одного исходного изображения сгенерировать множество уникальных обучающих примеров, что существенно повышает обобщающую способность модели. Обобщённая схема процесса аугментации данных для арктических условий представлена на рисунке 4.

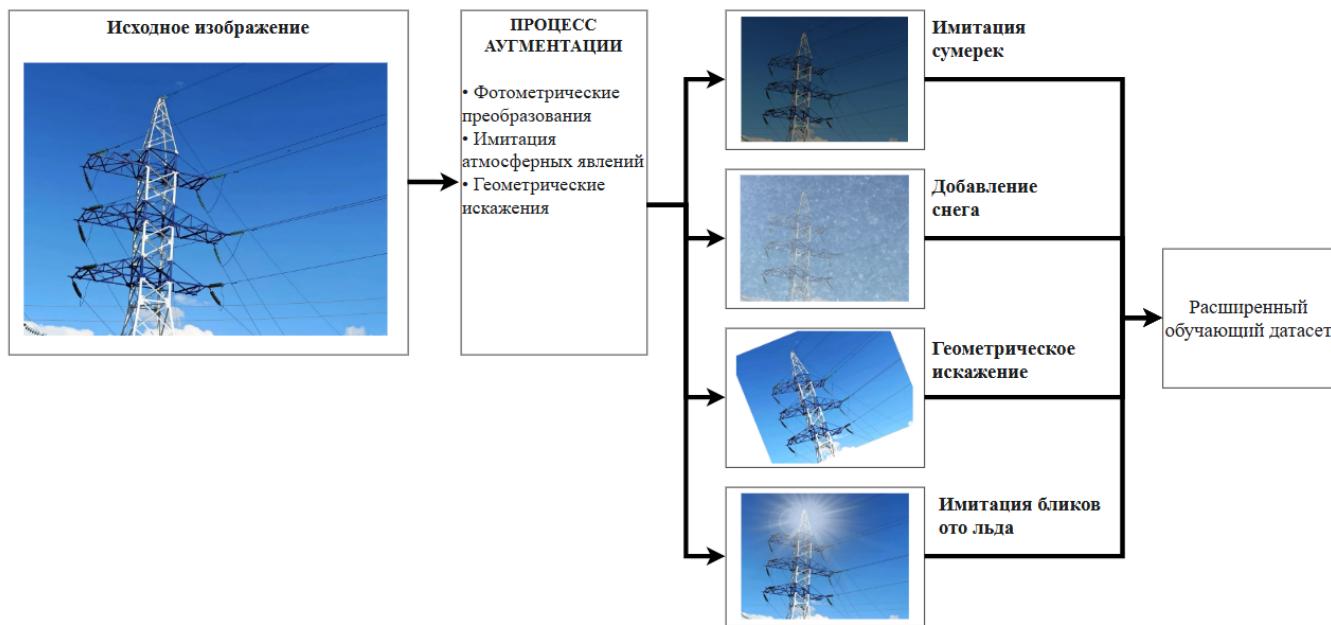


Рис. 4. Обобщённая схема процесса аугментации данных

Для решения комплексных задач диагностики, таких как обнаружение локальных перегревов, теоретической основой является мультимодальный анализ. Данный подход предполагает совместную обработку данных, полученных от сенсоров разной физической природы, как правило, из видимого (RGB) и длинноволнового инфракрасного диапазонов [14]. Архитектуры, реализующие данный принцип (Fusion CNN), объединяют признаки из обоих потоков данных на ранних или поздних слоях сети. Такой синтез информации позволяет одновременно идентифицировать конструктивный элемент объекта (по RGB-изображению) и его аномальное температурное состояние, что радикально снижает вероятность ложноположительных срабатываний по сравнению с анализом данных только из одного спектрального диапазона.

ные для них эксплуатационные риски и релевантные методы автоматизированной диагностики на основе нейросетевых моделей.

На основе анализа эксплуатационных уязвимостей объекты энергетической инфраструктуры были классифицированы по степени приоритетности для внедрения систем видеомониторинга. К группе с наивысшим приоритетом отнесены линейно-протяжённые объекты (линии электропередачи, трубопроводы), уязвимость которых определяется большой протяжённостью, труднодоступностью и прямым воздействием геокриологических и климатических факторов (деградация мерзлоты, ветровые и ледовые нагрузки). В группу со средним приоритетом включены узловые объекты (трансформаторные подстанции, ДЭС, котельные), риски которых связаны с износом оборудования, отказами отдельных агрегатов и необходимостью контроля технологических процессов.

Для оценки преимуществ автоматизированных систем выполнено сопоставление нейросетевого видеомониторинга с традиционными методами контроля, результаты которого представлены в Таблице 2.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Результаты проведённого аналитического исследования представляют собой систематизированную структуру данных, связывающую типы объектов энергетической инфраструктуры Арктики, характер-

Таблица 2

Сравнительный анализ традиционных и нейросетевых методов контроля

| Параметр сравнения | Традиционные методы контроля | Нейросетевой видеомониторинг |
|----------------------------------|---|--|
| Периодичность | Дискретная (плановые или аварийные осмотры) | Непрерывная или квазинепрерывная |
| Объективность | Зависимость от квалификации и состояния оператора | Формализованные, воспроизведимые критерии оценки |
| Пространственный охват | Локальный, ограниченный физическим доступом | Потенциально полный охват контролируемой зоны |
| Прогностический потенциал | Низкий (фиксация существующих дефектов) | Высокий (трекинг динамики развития дефектов) |
| Эксплуатационные затраты | Высокие (транспорт, персонал, спецтехника) | Низкие (после внедрения), автоматизированный режим |
| Зависимость от условий | Высокая (ограничения по погоде, времени суток) | Сниженная (при использовании IR-сенсоров и спец. оптики) |

Анализ показывает, что внедрение нейросетевых систем позволяет осуществить переход от реактивной модели обслуживания (устранение последствий) к проактивной, основанной на непрерывной оценке состояния и прогнозировании отказов.

Ключевым результатом исследования является разработанная матрица соответствия, представленная в таблице 1, которая систематизирует выбор технических решений для диагностики. Данная матрица устанавливает прямую связь между конкретным типом дефекта, задачей видеоанализа и оптимальной архитектурой нейронной сети. Выбор однопроходных архитектур (YOLO) для задач детектирования на протяжённых объектах обусловлен необходимостью обеспечения высокой скорости

обработки данных, получаемых, например, с БВС. Для количественной оценки площадных дефектов, таких как коррозия или обледенение, где точность определения границ является критичной, предпочтение отдаётся сегментационным моделям (U-Net, Mask R-CNN), обеспечивающим попиксельную классификацию [15].

Для обобщения представленных результатов и формирования прикладного инструментария разработана концептуальная модель принятия решений по выбору системы мониторинга, представленная на рисунке 5. Данная модель иллюстрирует последовательность шагов от идентификации объекта до определения конкретных технических требований к системе сбора и анализа данных.

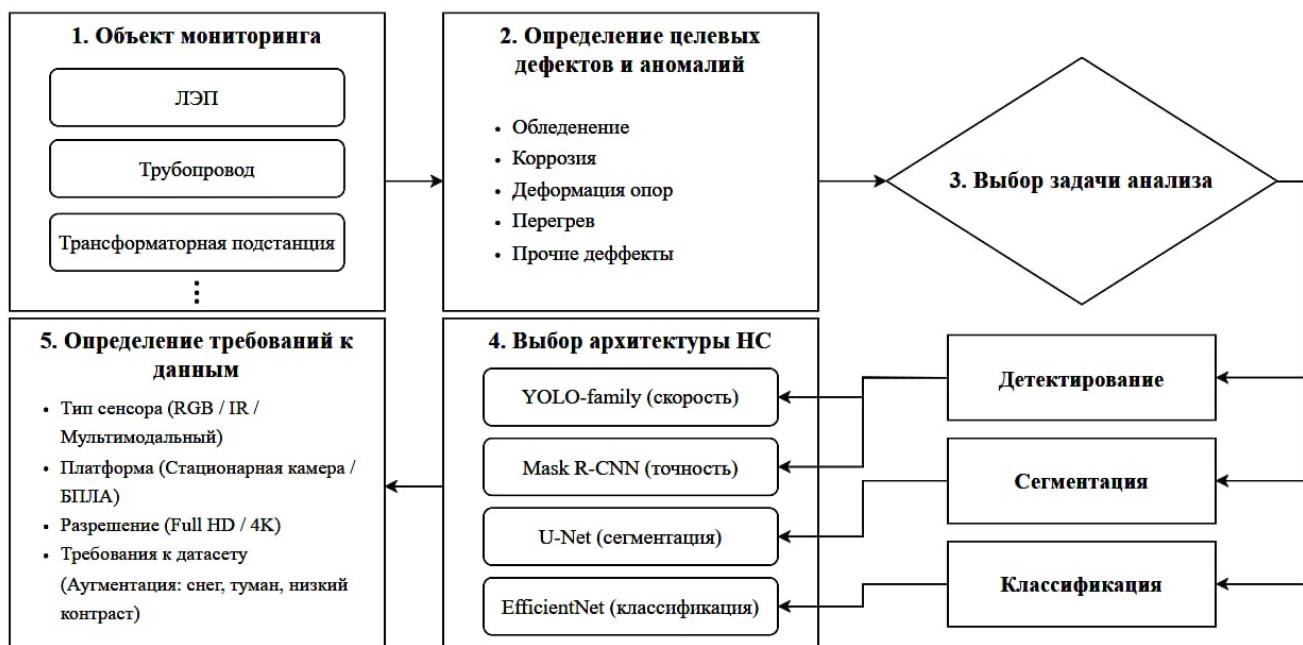


Рис. 5. Концептуальная модель выбора компонентов системы автоматизированного видеомониторинга

Модель демонстрирует, что для каждого класса объектов и характерных для него дефектов существует оптимальная комбинация метода анализа, нейросетевой архитектуры и способа получения визуальных данных. Такой подход позволяет унифицировать процесс проектирования систем автоматизированной диагностики и адаптировать их под специфические условия и задачи, существующие на объектах энергетической инфраструктуры в Арктической зоне РФ.

АНАЛИЗ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ

Представленные результаты свидетельствуют о высоком потенциале применения нейросетевых моделей видеомониторинга для автоматизации процессов технической диагностики энергетической инфраструктуры в Арктической зоне. Переход от дискретных инспекций к непрерывному автоматизированному контролю позволяет сформировать качественно новую модель эксплуатации, основанную на проактивном управлении состоянием активов и прогнозировании отказов. Однако реализация данного потенциала сопряжена с рядом фундаментальных технических и организационных барьеров, требующих детального рассмотрения.

Основным ограничением является дефицит репрезентативных обучающих датасетов. Эффективность любой нейросетевой модели прямо пропорциональна объёму и разнообразию данных, на которых она обучалась. Сбор и разметка достаточного количества изображений дефектов в арктических условиях представляют собой сложную и ресурсоёмкую задачу. Решение данной проблемы видится в комбинированном подходе, сочетающем использование техник аугментации, генерацию синтетических данных и организацию межотраслевого взаимодействия для создания единого национального репозитория визуальных данных по объектам промышленной инфраструктуры.

Вторым существенным барьером являются аппаратные и инфраструктурные ограничения. Эксплуатация систем видеомониторинга в условиях низких температур требует применения специализированного оборудования, включая термокожухи с системами подогрева для камер и вычислительных модулей. Данные системы повышают общее энергопотребление, что создаёт проблему автономного питания в удалённых районах, где доступ к централизованным сетям отсутствует, а эффективность солнечных панелей в период полярной ночи нулевая. Следовательно, возникает необходимость в разработке и применении энергоэффективных

вычислительных модулей (edge-computing devices) и оптимизированных, легковесных нейросетевых моделей, способных выполнять анализ непосредственно на объекте.

Проблема ограниченной пропускной способности каналов связи в арктических регионах диктует безальтернативность перехода к периферийным вычислениям. Передача необработанного видеопотока высокого разрешения для анализа в облачных сервисах в большинстве случаев технически невозможна или экономически нецелесообразна. Система мониторинга должна функционировать в автономном режиме, выполняя анализ на месте и передавая по узким каналам связи только структурированную метаинформацию о выявленных дефектах (тип, координаты, степень опасности, временная метка).

Перспективы развития данных технологий лежат в плоскости их интеграции с более сложными элементами систем управления, в частности с цифровыми двойниками энергетических объектов. Поступающие в реальном времени данные визуального и тепловизионного мониторинга, сегментированные и классифицированные нейросетевыми алгоритмами, становятся не просто потоком сырых фактов, а источником структурированных событий для верификации и калибровки физических моделей, заложенных в цифровой двойник. Такой симбиоз позволяет перейти от диагностики текущего состояния к предиктивному моделированию, существенно повышая точность прогнозирования остаточного ресурса критических компонентов и формируя оптимизированные, экономически обоснованные графики ремонтов и технического обслуживания.

Дальнейшая эволюция связана с созданием гибридных моделей, где данные компьютерного зрения ищут корреляции с параметрами технологических режимов, регистрируемых SCADA-системами. Данный подход открывает путь к решению обратных задач, например, определению глубины коррозии по динамике тепловых полей или оценке механической прочности опоры ЛЭП по видеоданным о её вибрации в совокупности с метеорологической информацией.

В аппаратном аспекте ключевым направлением является развитие отечественной компонентной базы для edge-устройств и сенсоров, что представляет собой стратегическую задачу обеспечения технологического суверенитета. Речь идет не только об импортозамещении, но и о создании специализированных решений, архитектурно оптимизированных для работы в условиях Арктики. Это включает разработку энергоэффективных процессоров с

аппаратными ускорителями для нейросетевых вычислений, устойчивых к длительному воздействию низких температур, а также мультиспектральных камерных модулей, изначально спроектированных для работы в условиях слепящих бликов от снега и низкой освещенности полярной ночи.

Успех в этой области позволит создать замкнутый технологический цикл — от сбора данных до принятия автономных решений на основе отечественных платформ, что является залогом устойчивого и безопасного развития арктической энергетики.

ЗАКЛЮЧЕНИЕ

Нейросетевые технологии видеомониторинга являются перспективным инструментом для автоматизации контроля и диагностики технического состояния энергетической инфраструктуры в Арктической зоне РФ. Их применение позволяет преодолеть ограничения традиционных методов, связанные с дискретностью, субъективностью и высокими эксплуатационными затратами.

Успешная реализация систем автоматизированной диагностики требует системного подхода, включающего выбор архитектуры нейронной сети, релевантной поставленной задаче (детектирова-

ние, сегментация, классификация), и формирование специализированных обучающих датасетов с применением техник аугментации, адаптированных под специфику арктических визуальных сцен.

Ключевыми техническими барьерами для широкого внедрения данных систем являются аппаратные ограничения (энергоавтономность, устойчивость к низким температурам), дефицит размеченных данных и низкая пропускная способность каналов связи, что обуславливает необходимость применения технологий периферийных вычислений.

Внедрение автоматизированных систем видеомониторинга способствует повышению надёжности и безопасности функционирования энергетических объектов, снижению вероятности аварийных ситуаций и оптимизации процессов эксплуатации, что вносит вклад в обеспечение устойчивого социально-экономического развития северных территорий Российской Федерации.

Направлениями для дальнейших исследований являются разработка унифицированных протоколов сбора и передачи диагностических данных, создание мультимодальных моделей, интегрирующих информацию из различных спектральных диапазонов (RGB+IR), и формирование отечественной базы эталонных изображений дефектов арктической инфраструктуры.

СПИСОК ЛИТЕРАТУРЫ

1. Соловьев Д. Климатические изменения и энергетическая инфраструктура в Российской Арктике // Проект Байкал. – 2022. – Т. 19, № 71. – С. 50-57. – DOI 10.51461/projectbaikal.71.1941. – EDN CWWXKD.
2. Биев А. А. Формирование территориальной инфраструктуры обеспечения топливно-энергетическими ресурсами в Арктической зоне России // Север и рынок: формирование экономического порядка. – 2019. – № 3(65). – С. 43-51. – DOI: 10.25702/KSC.2220-802X.2019.65.3.43-51. – EDN: SKSJHE.
3. Бирюков Н.А., Бирюков Ю.А., Кравченко И.Н., Добрышкін Е.О., Петров Г.Д., Гончаров Н.О. Повышение эффективности и обеспечение энергетической безопасности арктической инфраструктуры с применением программного обеспечения // Строительные и дорожные машины. – 2020. – № 8. – С. 51-56. – EDN: GOYMAC.
4. Шабунин А. С., Чернецкий М. Ю., Осиповский Р. В. Нейросетевые модели технологического оборудования в системе мониторинга и предиктивной аналитики // Электрические станции. – 2023. – № 11(1108). – С. 50-57. – DOI: 10.34831/EP.2023.1108.11.007. – EDN: BPMKTG.
5. Голышев М. Н. Современные методы диагностики и контроля высоковольтного оборудования подстанции // Сборник статей V Всероссийской научно-практической конференции молодых ученых и студентов, Пенза, 30 июня 2023 года / Под научной редакцией О.Н. Кухарева, А.В. Носова. – Пенза: Пензенский государственный аграрный университет, 2023. – С. 107-110. – EDN: RNSLOJ.
6. Terven Ju., Córdova-Esparza D. M., Romero-González, Ju. A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS // Machine Learning and Knowledge Extraction. – 2023. – Vol. 5, No. 4. – P. 1680-1716. – DOI: 10.3390/make5040083. – EDN: XZSDRH.
7. Chen H., Zhang Zh. Hybrid neural network based on novel audio feature for vehicle type identification // 2020 IEEE Intern. Instrumentation and Measurement Technology Conf. (I2MTC). 25-28 May 2020. DOI: 10.1109/I2MTC43012.2020.9129183

8. Скобцов В. Ю., Соколов Б. В., Чжан В. А., Фу М. Гибридные нейросетевые модели мониторинга данных временных рядов сложных объектов // Известия высших учебных заведений. Приборостроение. – 2024. – Т. 67, № 2. – С. 200-204. – DOI: 10.17586/0021-3454-2024-67-2-200-204. – EDN: BSJAIK.
9. Wang S., Zhao J., Ta N., et al. 2021. A real-time deep learning forest fire monitoring algorithm based on an improved Pruned+KD model. Journal of Real-Time Image Processing 18 (6): 2319–2329. DOI: 10.1007/s11554-021-01124-9. – EDN: NSIREW
10. Siddique N., Paheding S., Elkin C. P., Devabhaktuni V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications // IEEE Access. – 2021. – Vol. 9. – P. 82031-82057. – DOI: 10.1109/ access.2021.3086020. – EDN: CHQTJF.
11. Biibosunova S. K., A. Asylbek Kyzy, Duyshembieva B. D., Kojoev K. T. Artificial intelligence methods and neural networks for solving forecasting problems // Bulletin Kyrgyz State University named after I. Arabaev. – 2024. – No. 3-2. – P. 24-30. – DOI: 10.33514/1694-7851-2024-3/2-24-30. – EDN: MAFCLP.
12. Игнатьева С. А., Богуш Р. П. Увеличение точности реидентификации людей на основе двухэтапного обучения сверточных нейронных сетей и аугментации // Информатика. – 2023. – Т. 20, № 1. – С. 40-54. – DOI 10.37661/1816-0301-2023-20-1-40-54. – EDN: UKBNNG.
13. Fu D. Unsupervised pre-training for person re-identification // 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20-25 June 2021.- Nashville, 2021.- P. 14745-14754. DOI: 10.1109/ CVPR46437.2021.01451
14. Конарев Д. И. Гуламов А. А. Повышение точности предварительно обученных нейронных сетей путем тонкой настройки // Материалы конф. "Информационные технологии в управлении", Санкт-Петербург, 6-8 окт. 2020 г.- СПб., 2020.- С. 200-212. EDN: SDKTPH
15. Xie T. Cut-thumbnail: A novel data augmentation for convolutional neural network // Proc. of the 29th ACM Intern. Conf. on Multimedia, Virtual Event, China, 20-24 Oct. 2021.- Virtual Event, China, 2021.- P. 1627-1635.

Аппаратный ускоритель вывода глубоких нейронных сетей на основе СФ-блоков блочной ОЗУ со встроенными вычислительными ядрами

A.A. Spiridonov, O.V. Zobov

Hardware Accelerator for Deep Neural Network Inference Based on SF Blocks of Block RAM with Integrated Computing Cores

Abstract. This paper explores the development of a hardware accelerator architecture for deep neural network inference on programmable logic integrated circuits. This architecture utilizes complex functional blocks of block RAM with embedded computational cores. The proposed approach implements an in-memory computing paradigm and aims to reduce the amount of data exchange between memory and computational logic, mitigating the impact of the "memory wall" when processing low-bit data. It is demonstrated that the architecture supports variable-bit quantization while maintaining standard access via one port of the dual-port memory, enabling joint operation with digital signal processing units. A hybrid computing system is described in which the proposed blocks perform bulk multiply-accumulate operations, while the digital signal processing units provide high-precision operations and result fusion. Simulation results are presented, confirming an increase in peak performance compared to existing in-memory computing architectures for PLIs.

Keywords: deep learning, hardware accelerators, in-memory computing, variable-bit quantization, transformer neural networks, programmable logic integrated circuits, block RAM.

подтверждающие повышение пиковой производительности по сравнению с существующими архитектурами вычислений в памяти для программируемых логических интегральных схем.

Ключевые слова: глубокое обучение, аппаратные ускорители, вычисления в памяти, квантование с переменной разрядностью, трансформерные нейронные сети, программируемые логические интегральные схемы, блочная оперативная память.

ВВЕДЕНИЕ

Глубокие нейронные сети (ГНС) демонстрируют впечатляющие результаты в широком спектре прикладных задач [1, 2], включая классификацию объектов на периферийных устройствах [3] и предиктивный анализ [4], находя всё более широкое применение в цифровых информационно-управляющих системах [5, 6], в том числе для оптимизации вывода нейронных сетей в автономной робототехнике [7] и автоматизации промышленных процессов [8].

Особенно актуальным становится их использование в контексте автономных мобильных роботов (AMP), где требуется анализ потоков данных с разнообразных сенсоров в реальном времени для решения задач навигации, распознавания динамического окружения и принятия управляющих решений в условиях ограниченных вычислительных и энергетических ресурсов [9].

Однако экспоненциальный рост сложности современных ГНС, таких как многослойные свёрточные сети и трансформерные архитектуры, а также объёмы обрабатываемых ими данных (например, видеопотоки высокого разрешения), предъявляют

O.V. Зобов¹A.A. Спиридоносов²

¹Главный инженер по робототехнике АО «НПК «Криптонит».

E-mail: o.zobov@kryptonite.ru
RSCL: 7699-7709

²Руководитель лаборатории информационной и сетевой безопасности АО «НПК «Криптонит».
a.spiridonov@kryptonite.ru

Аннотация. Статья посвящена разработке архитектуры аппаратного ускорителя вывода глубоких нейронных сетей на программируемых логических интегральных схемах, основанной на сложных функциональных блоках блочной оперативной памяти со встроенными вычислительными ядрами. Предложенный подход реализует парадигму вычислений в памяти и направлен на снижение объёма обмена данными между памятью и вычислительной логикой, что смягчает влияние «стены памяти» при обработке низкоразрядных данных. Показано, что архитектура поддерживает квантование с переменной разрядностью при сохранении стандартного доступа по одному из портов двухпортовой памяти, что обеспечивает совместную работу с блоками цифровой обработки сигналов. Описана гибридная организация вычислительной системы, в которой предлагаемые блоки выполняют массовые операции умножения-накопления, а блоки цифровой обработки сигналов обеспечивают операции повышенной точности и объединение результатов. Приведены результаты моделирования, подтверждающие повышение пиковой производительности по сравнению с существующими архитектурами вычислений в памяти для программируемых логических интегральных схем.

всё более высокие требования к производительности и энергоэффективности аппаратных ускорителей. При этом исследование архитектурных и схемотехнических основ таких устройств ведётся на протяжении многих лет [10], а постоянное развитие алгоритмов ГНС стимулирует поиск новых аппаратных решений.

Для аппаратного ускорения ГНС используются различные платформы, включая центральные процессоры, графические ускорители и специализированные интегральные схемы. В этом ряду программируемые логические интегральные схемы (ПЛИС) занимают особую нишу благодаря своей реконфигурируемости, позволяющей создавать аппаратные архитектуры, точно соответствующие структуре конкретной нейронной сети, и возможности организации широкого параллелизма по данным [11]. Это делает их привлекательной платформой для создания структурно-фиксированных ускорителей ГНС, особенно для приложений, где важны низкие задержки, энергоэффективность и возможность адаптации аппаратной части под различные алгоритмы.

Одним из ключевых методов улучшения производительности и энергоэффективности ГНС на ПЛИС, как и на других платформах, является квантование, позволяющее снизить вычислительную сложность, требования к пропускной способности памяти и объёму хранения весовых коэффициентов за счёт использования данных пониженной разрядности.

СОВРЕМЕННЫЕ ПОДХОДЫ К АППАРАТНОМУ УСКОРЕНИЮ КВАНТОВАННЫХ ГНС НА ПЛИС

Ведётся активная разработка ПЛИС-ускорителей для квантованных ГНС, направленная на достижение низких задержек и высокой пропускной способности в задачах распознавания образов, в том числе для промышленных приложений [6, 8]. При этом применяются разнообразные архитектурные решения и методы оптимизации, адаптированные под конкретные задачи и аппаратные платформы.

Например, для задач семантической сегментации, критически важных для систем компьютерного зрения автономных транспортных средств, в работе [12] предлагается конвейерная потоковая архитектура для сети типа U-Net, развёрнутая на ПЛИС Xilinx Alveo U250. Ключевой особенностью является использование 4-битного целочисленного квантования, что позволяет существенно снизить требования к аппаратным ресурсам и одновременно достичь производительности обработки около 23 изображе-

ний в секунду с задержкой 44 мс на наборе данных Cityscapes при среднем пересечении по объединению 62,9%. Авторы подчёркивают, что выбранный подход, ориентированный на высокую степень параллелизма и низкую задержку, выгодно отличается от более распространённых рекурсивных архитектур на основе специализированных процессорных модулей, которые, будучи эффективными по ресурсам, обычно характеризуются более высокими задержками.

Усилия по минимизации задержек, являющихся основной проблемой из-за высокой вычислительной сложности ГНС, также демонстрируются в [13], где для ускорителя сети YOLOv2-tiny (с входным разрешением 1280×384), предназначенного для детекции объектов в реальном времени, используется ПЛИС-СНК Xilinx ZC706. В этой работе представлена мелкозернистая конвейерная архитектура с поколоночной обработкой (последовательной обработкой столбцов входного тензора), спроектированная для минимизации времени запуска конвейера. Дополнительное сокращение задержек достигается за счёт техники пропуска операций заполнения, использования корректирующей схемы для операций знакового умножения и модуля пулинга с общим буфером для снижения затрат на хранение данных. В результате этих решений удалось уменьшить задержку обработки в 2,125–2,34 раза по сравнению с предшествующими ПЛИС-ускорителями для YOLOv2-tiny, что критично для встраиваемых и автономных систем.

В условиях эксплуатации, характеризующихся высокой алгоритмической и архитектурной сложностью, например в многозадачных встраиваемых робототехнических системах, где одновременный доступ нескольких процессов к ускорителю ГНС может приводить к конфликтам аппаратных ресурсов, предлагаются специализированные решения.

В работе [14] описывается ускоритель с поддержкой прерываний INCA (Interruptible CNN Accelerator), реализованный на встраиваемой платформе на базе ПЛИС-СНК. Ключевой особенностью является метод прерывания на основе виртуальных инструкций, позволяющий приостанавливать и возобновлять любую выполняемую задачу ГНС. Это обеспечивает одновременный доступ к ускорителю от нескольких процессов, например, для извлечения признаков и распознавания локации при распределённом SLAM-алгоритме, со значительно сниженной (до 1% от базового уровня) задержкой отклика на прерывание по сравнению с прерыванием на границах слоёв [14], повышая тем самым отзывчивость и надёжность многозадачных робототехнических систем.

Стоит отметить, что во всех упомянутых работах оценки производительности проводились на стандартных наборах данных (например, Cityscapes) и не всегда обсуждается развертывание в реальных полевых условиях эксплуатации робототехнических систем [12]. Тем не менее эти исследования подтверждают эффективность применения ПЛИС и квантования с фиксированной низкой точностью для ускорения ГНС [11]. Однако для дальнейшего повышения эффективности и гибкости перспективным направлением является квантование с переменной разрядностью, при котором различные компоненты сети (веса и активации на разных слоях или даже внутри одного слоя) могут иметь разную битовую ширину (например, от 2 до 8 бит). Это позволяет более тонко находить компромисс между точностью модели, её размером и вычислительными затратами.

Несмотря на очевидные преимущества, реализация эффективных и гибких ускорителей ГНС с переменной разрядностью на ПЛИС сталкивается с рядом существенных проблем, не всегда полностью решаемых вышеописанными подходами. Эти проблемы обусловлены как архитектурными особенностями самих ПЛИС, так и ограничениями традиционных подходов к проектированию вычислительных блоков и организации потоков данных.

Постановка задачи и ограничения существующих подходов

Одной из ключевых проблем является неэффективное использование стандартных вычислительных ресурсов ПЛИС, таких как блоки цифровой обработки сигналов (ЦОС), при работе с данными переменной и пониженной точности. Блоки ЦОС разработаны для операций с фиксированной, относительно высокой разрядностью операндов. Применение их для обработки данных, квантованных до 2–8 бит, без специализированных архитектурных решений приводит к существенной недозагрузке аппаратных вычислителей, что снижает общую эффективность ускорителя в терминах производительности на единицу площади кристалла и потребляемой мощности. Хотя существуют методики, такие как уплотнение операций в блоках ЦОС, их универсальность и выигрыш в производительности сильно зависят от конкретных комбинаций точности весов и активаций, а также от внутренней архитектуры самого блока ЦОС, не позволяя полностью раскрыть потенциал низкоразрядных вычислений.

Другой комплекс проблем связан с ограничениями существующих архитектур вычислений в памяти, которые предлагались для ПЛИС как способ повы-

шения производительности. Ранние подходы часто требовали специфической организации данных в блоках оперативной памяти (ОЗУ), например, транспонированного размещения, что несовместимо со стандартным доступом к данным. Более того, такие архитектуры могли полностью монополизировать порты доступа к блоку ОЗУ во время вычислительных операций, исключая возможность параллельной работы с другими вычислительными ресурсами ПЛИС, такими как ЦОС, и затрудняя реализацию критически важных для производительности техник, например, двойной буферизации. Последующие разработки, хотя и решали некоторые из этих проблем, могли иметь ограничения в гибкости поддержки переменной точности или предлагали фиксированные схемы параллелизма, что не позволяло эффективно адаптироваться к разнообразию топологий современных ГНС.

Значительные издержки по времени и энергии, связанные с передачей больших объёмов весовых коэффициентов и активаций между модулями памяти и удалёнными вычислительными блоками, также остаются серьёзным препятствием на пути к созданию высокопроизводительных и энергоэффективных ускорителей. Это явление, известное как «стена памяти» — несоответствие темпов роста пропускной способности памяти росту вычислительной производительности — ограничивает масштабируемость традиционных архитектур, где память и логика физически разделены [14]. Сокращение этих перемещений данных является одной из приоритетных задач при проектировании аппаратных ускорителей ГНС.

Наконец, по мере роста сложности и размеров ГНС количество доступных на кристалле ПЛИС высокопроизводительных вычислительных ресурсов, таких как блоки ЦОС, может стать ограничивающим фактором. Для дальнейшего наращивания вычислительной мощности требуется либо увеличение числа таких специализированных блоков, что ведёт к росту стоимости и энергопотребления кристалла, либо поиск альтернативных подходов к распределённым вычислениям, использующих другие доступные на ПЛИС ресурсы, что также перекликается с общими тенденциями развития высокопроизводительных вычислительных платформ [16].

Актуальность, научная новизна и практическая значимость исследования

Актуальность настоящего исследования определяется критическими ограничениями современных аппаратных ускорителей глубоких нейронных сетей (ГНС) на программируемых логических интеграль-

ных схемах (ПЛИС) при обработке квантованных данных. Три взаимосвязанные проблемы препятствуют эффективному применению методов квантования: недоиспользование высокоразрядных умножителей блоков цифровой обработки сигналов (ЦОС) при низкоразрядных операциях; архитектурные ограничения существующих решений вычислений в памяти, не позволяющие гибко сочетать их с традиционными вычислительными ресурсами; возрастающее влияние «стены памяти» (диспропорции между пропускной способностью памяти и вычислительной производительностью).

Стремительное развитие ГНС и их внедрение в критические области (автономная навигация, распознавание образов в реальном времени, робототехника) требует создания гибких аппаратных платформ, способных адаптироваться к разнообразным стратегиям квантования. ПЛИС, обладая реконфигурируемостью и битовой программируемостью, представляют перспективную основу для таких решений. Однако существующие подходы либо ограничивают гибкость квантования фиксированной разрядностью, либо монополизируют ресурсы блочной памяти, исключая эффективное взаимодействие с блоками ЦОС. Необходим качественно новый подход, сочетающий вычисления в памяти с традиционными вычислительными ресурсами ПЛИС.

Цель работы состоит в разработке архитектуры аппаратного ускорителя вывода ГНС на ПЛИС, основанной на сложных функциональных блоках (СФ-блоках), которые интегрируют блочную оперативную память со встроенными вычислительными ядрами. Ключевая особенность архитектуры заключается в сохранении одного порта двухпортовой памяти доступным для стандартных операций, что обеспечивает беспрепятственное взаимодействие с блоками ЦОС и возможность двойной буферизации. Предлагаемое решение реализует парадигму вычислений в памяти: операции умножения-накопления выполняются непосредственно у данных, минимизируя объём передач между памятью и логикой.

Архитектура нативно поддерживает переменную точность (весовые коэффициенты разрядностью 2, 4 или 8 бит; активации от 2 до 8 бит) с линейным масштабированием производительности, устранив ограничения стандартных блоков ЦОС. Система построена по гибридному принципу: СФ-блоки выполняют массовые низкоразрядные операции, в то время как блоки ЦОС обеспечивают высокоточные вычисления и объединение результатов.

Методология исследования включает несколько взаимодополняющих подходов. Архитектура

разрабатывалась на уровне регистровых передач с последующим синтезом ключевых компонентов в среде Xilinx Vivado Design Suite. Аппаратные затраты (площадь кристалла, задержки сигналов, энергопотребление) моделировались инструментом COFFE с использованием 22-нанометровой предиктивной технологической модели, репрезентативной для устройств класса Xilinx UltraScale+ (технологический узел 16 нм). Пиковая производительность системы оценивалась экстраполяцией результатов синтеза на эталонные ПЛИС семейств Xilinx Virtex UltraScale+.

Количественное сравнение с существующими архитектурами вычислений в памяти (CCB) и модифицированными блоками ЦОС (eDSP) проводилось на базе опубликованных данных и собственного моделирования с точностью до такта. Оценка выполнялась для эталонных моделей ГНС (AlexNet, VGG-16, ResNet-18, ResNet-34, Vision Transformer) на задаче классификации ImageNet.

Научная новизна работы определяется совокупностью оригинальных архитектурных решений. Впервые предложена архитектура СФ-блоков на базе блочной памяти ПЛИС, сохраняющая стандартный доступ по одному порту во время выполнения вычислений. Это позволяет блокам ЦОС свободно обращаться к данным и реализовывать двойную буферизацию без внешних буферов, в отличие от предшествующих архитектур вычислений в памяти, полностью монополизирующих порты доступа.

Разработана оригинальная структура вычислительного ядра, сочетающая разрядно-последовательную обработку активаций с битово-параллельной обработкой весовых коэффициентов. Такая организация обеспечивает гибкую поддержку переменной точности: веса могут иметь разрядность 2, 4 или 8 бит; активации варьируются от 2 до 8 бит. При этом задержка операций масштабируется линейно в зависимости от разрядности активаций.

Предложена схема дублирования данных внутри блочной памяти, эксплуатирующая вычислительный параллелизм ГНС двумя способами: через совместное использование весовых коэффициентов (коэффициент N принимает значения 1, 2 или 4) и через совместное использование активаций (параметр N_w достигает 16 для 2-битных весов). Это существенно повышает эффективность использования аппаратуры по сравнению с архитектурами, поддерживающими только один тип параллелизма.

Обоснована концепция гибридной вычислительной системы, в которой СФ-блоки и блоки ЦОС функционируют как комплементарные вычислительные движки с общим кэшем данных. Такая организация

обеспечивает гибкое распределение нагрузки в зависимости от разрядности слоёв ГНС.

Практическая значимость заключается в существенном расширении возможностей ПЛИС для ускорения современных ГНС с переменной разрядностью. Моделирование демонстрирует пиковую пропускную способность до 50,7 ТОП/с для 2-битных операций и 6,9 ТОП/с для 8-битных операций. Это превышает производительность существующих архитектур вычислений в памяти в 2,7 раза и 1,7 раза соответственно. Экспериментальная оценка на эталонных моделях показывает среднее ускорение в 2,16 раза при квантовании активаций до 6 бит с потерей точности менее 0,5 % на задаче классификации ImageNet.

Архитектура обеспечивает более эффективное масштабирование производительности на единицу площади кристалла по сравнению с блоками ЦОС. Для низкоразрядных операций выигрыш составляет от 2 до 3 раз. Разработанные решения могут применяться в специализированных вычислительных системах для компьютерного зрения, обработки естественного языка и других приложений, требующих интенсивных вычислений с низкоразрядными данными. Целевые платформы включают встраиваемые системы с ограниченными энергетическими ресурсами: автономные мобильные роботы, серверы локальной обработки данных, системы реального времени.

ПРЕДЛАГАЕМАЯ АРХИТЕКТУРА

Для преодоления указанных ограничений и повышения эффективности ПЛИС-ускорителей ГНС разработана архитектура СФ-блока на базе блочной ОЗУ с поддержкой обработки данных разнородной точности, целенаправленно решающая проблемы недоиспользования ресурсов и «стены памяти». Предлагаемое решение базируется на принципе вычислений в памяти: вычислительная логика располагается непосредственно у данных в блочной ОЗУ. Это позволяет выполнять часть операций без выгрузки весов и активаций во внешние вычислители, тем самым уменьшает объём и частоту передач между памятью и логикой и, как следствие, снижает задержки и повышает суммарную производительность. Дополнительно обеспечивается нативная поддержка квантования с переменной разрядностью — слабое место стандартных блоков цифровой обработки сигналов. Разрабатываемый СФ-блок представляет собой модифицированный модуль блочной ОЗУ, способный автономно выполнять операции умножения-накопления, при этом

один из портов двухпортовой памяти сохраняется для стандартного доступа со стороны других вычислителей. Таким образом, решение не просто добавляет вычислительные ресурсы, а меняет парадигму взаимодействия памяти и логики в ПЛИС-ускорителях ГНС.

Организация гибридной вычислительной системы

Организация аппаратного ускорителя ГНС с использованием блочной ОЗУ со встроенными вычислительными ядрами и традиционных блоков цифровой обработки сигналов предполагает создание гибридной вычислительной системы, где вычислительные ресурсы различного типа интегрируются для достижения максимальной общей производительности и энергоэффективности ускорителя. В такой системе каждый компонент специализируется на выполнении тех операций или обработке тех типов данных, для которых он спроектирован.

СФ-блоки с встроенными вычислительными ядрами наиболее эффективны для выполнения основной массы вычислительно интенсивных операций в ГНС, особенно там, где возможно применение квантования с переменной или пониженной точностью (например, 2–8 бит). Они подходят для высокопараллельных свёрточных и полно связанных слоёв, где способность к массовому выполнению операций умножения-накопления с низкоразрядными весами и активациями раскрывается в полной мере. За счёт выполнения вычислений непосредственно в памяти, где хранятся веса, минимизируются задержки на передачу данных, что критично для слоёв с большим количеством параметров или при интенсивном повторном использовании операндов.

Таким образом, на вычислительные ядра предлагаемой архитектуры целесообразно возлагать задачи обработки основной части свёрток и полно связанных слоёв в таких сетях, как AlexNet, VGG, ResNet, широко применяемых для распознавания визуальных образов [19], а также квантованные компоненты трансформеров, где приоритетом является высокая пропускная способность и энергоэффективность за счёт работы с данными пониженной точности. Их применение позволяет существенно разгрузить другие вычислительные ресурсы ПЛИС от рутинных, но объёмных вычислений.

Традиционные блоки цифровой обработки сигналов применяются в операциях, требующих высокой точности представления данных (например, 16-бит, 18-бит или даже эмуляции операций с плавающей точкой), которые могут быть критичны для сохранения общей точности модели, особенно в

первых или выходных слоях ГНС, а также в специфических задачах регрессии. Кроме того, они эффективно справляются с реализацией сложных математических функций, таких как нелинейные функции активации, операции нормализации или другие специализированные вычисления, если их аппаратная реализация в этих блоках более производительна, чем синтез на программируемой логике.

Также блоки цифровой обработки сигналов предпочтительны для обработки слоёв с нерегулярной структурой вычислений или низкой степенью внутреннего параллелизма, где трудно эффективно задействовать множество мелких идентичных ядер

предлагаемой архитектуры. Накопление промежуточных результатов с высокой точностью, чтобы избежать переполнения при суммировании большого числа низкоразрядных произведений, также является сильной стороной этих блоков благодаря их широким аппаратным аккумуляторам. Они могут использоваться для агрегации результатов, полученных от ядер предлагаемой архитектуры, и для финальной постобработки данных.

Структурная схема аппаратного ускорителя ГНС с использованием блочной ОЗУ с встроенными вычислительными ядрами и традиционных блоков ЦОС представлена на рисунке 1.

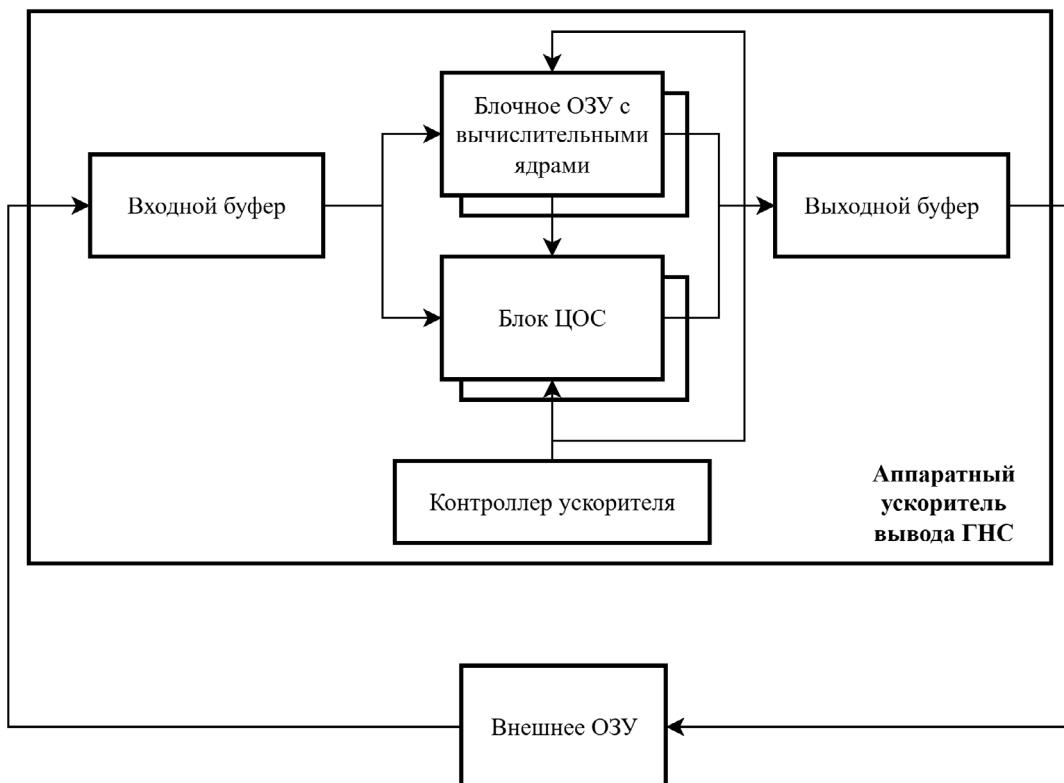


Рис. 1. Структурная схема аппаратного ускорителя вывода ГНС на базе блочных ОЗУ с встроенными вычислительными ядрами

Ускоритель состоит из следующих функциональных блоков:

1. Массив блоков предлагаемой архитектуры

служит как для хранения весовых коэффициентов и промежуточных данных ГНС, так и для выполнения значительной части операций умножения-накопления непосредственно в памяти, особенно для слоёв или частей модели, использующих квантование с переменной или пониженной точностью.

2. **Массив блоков цифровой обработки сигналов** используется для выполнения операций умножения-накопления, которые могут быть неэффективны или сложны для реализации в рамках вычислительных ядер предлагаемой архитектуры (например, операции с высокой точностью, спе-

цифические нелинейные функции, если они аппаратно реализованы в блоках цифровой обработки сигналов).

3. **Управляющая логика** (центральный контроллер) координирует работу всего ускорителя, включая:

- декомпозицию вычислительной нагрузки между блоками предлагаемой архитектуры и блоками ЦОС;

- генерацию и рассылку управляющих команд (инструкций вычислений в память, СИМ, для предлагаемой архитектуры, команд для блоков цифровой обработки сигналов);

- управление потоками данных между различными компонентами;

- синхронизацию стадий загрузки/выгрузки и вычислений в конвейере обработки блоков данных с двойной буферизацией: загрузку входов и фильтров, выполнение УНК в СФ-блоках (разрядно-последовательная обработка активаций) и в блоках ЦОС, чтение и обнуление частичных сумм и запись результатов, а также арбитраж доступа к двухпортовой блочной ОЗУ и согласование тактовых доменов при работе вычислительной логики в режиме удвоенной частоты относительно памяти.

4. **Интерфейс с внешней ОЗУ** для загрузки весов модели, входных данных и выгрузки конечных результатов.

5. **Внутренние буферы** данных могут быть реализованы на стандартных блоках ОЗУ или блоках предлагаемой архитектуры, работающих в режиме

обычной памяти, для временного хранения активаций, частичных сумм и других промежуточных результатов.

Принцип организации СФ-блока на базе блочной ОЗУ с встроенной вычислительной логикой

Принцип организации СФ-блока на базе блочной ОЗУ построен на минимальном изменении стандартного интерфейса доступа к памяти при существенном расширении её функциональности за счёт встроенной вычислительной логики. Цель — обеспечить обратную совместимость на уровне сигналов и простую интеграцию в проекты на ПЛИС-СНК, одновременно предоставляя инструмент ускорения вывода ГНС. Структурная схема СФ-блока приведена на рисунке 2.

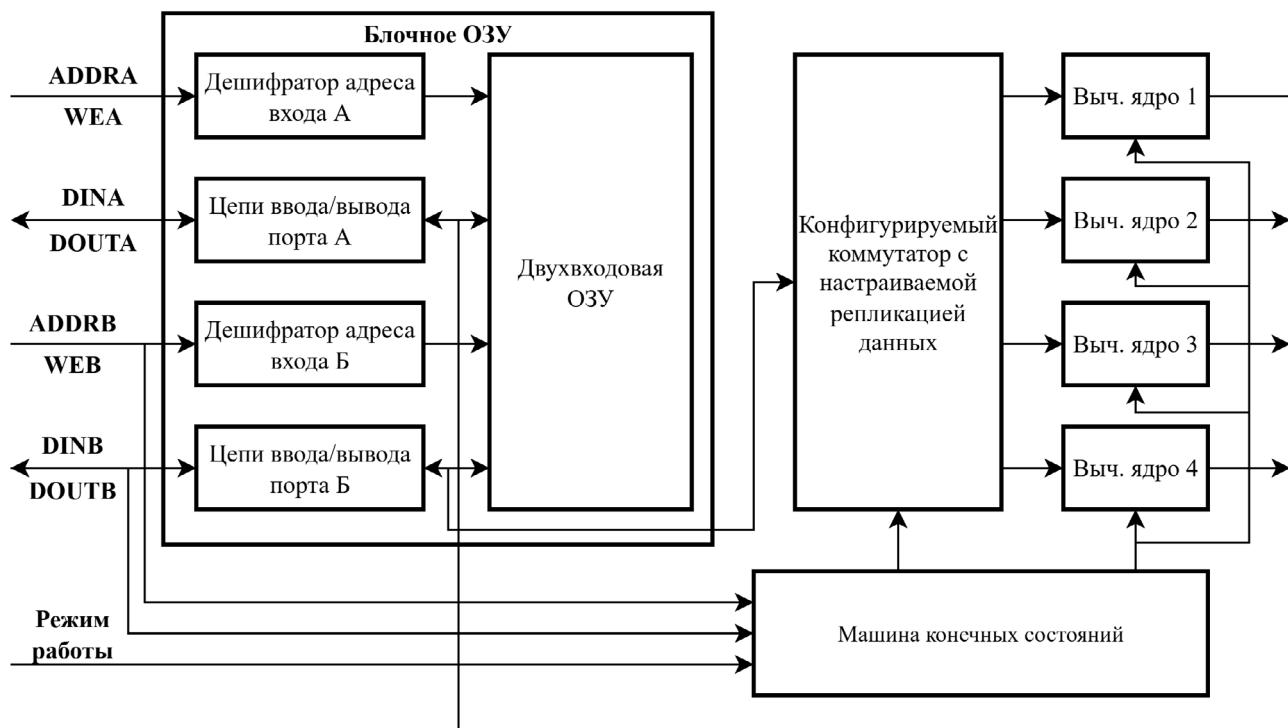


Рис. 2. Структурная схема блочной ОЗУ с встроенными вычислительными ядрами

СФ-блок на базе двухпортовой блочной ОЗУ сохраняет стандартный набор интерфейсных сигналов. Для каждого из двух портов (Порт А и Порт В) используются: тактовый сигнал; сигнал разрешения порта (ENARDEN, ENBWREN); адресные шины (ADDRA[15:0], ADDRB[15:0]); входные шины данных для записи (DINA[35:0], DINB[35:0]); выходные шины данных для чтения (DOUTA[35:0], DOUTB[35:0]); сигналы управления записью (WEA[3:0], WEBWE[3:0]).

В обычном режиме СФ-блок функционирует как стандартная двухпортовая память, что обеспечивает бесшовную интеграцию в существующие проекты, где он выполняет традиционные функции хранения данных. Переключение между режимом

стандартной памяти и режимом вычислений выполняется управляющей логикой СФ-блока (встроенным конечным автоматом): признак команды вычислений в памяти передаётся через зарезервированный диапазон адресов и/или специальную комбинацию управляющих сигналов, при этом семантика стандартных сигналов памяти для пользовательского порта не изменяется.

Встроенная вычислительная логика (вычислительные ядра) располагается в непосредственной близости к данным внутри СФ-блока и предназначена для выполнения операций умножения-накопления непосредственно у ячеек хранения. Концептуально она развивает идею использования

вспомогательных вычислительных массивов в блоках памяти, предоставляя расширенную поддержку квантования с переменной точностью и более гибкие механизмы управления параллелизмом.

Выполнение части операций без выгрузки весов и активаций во внешние вычислители уменьшает объём и частоту передач между памятью и логикой, что снижает задержки и повышает суммарную производительность; одновременно обеспечивается нативная поддержка переменной точности – слабое место стандартных блоков ЦОС. Параллельная организация встроенной вычислительной логики внутри одного СФ-блока обеспечивает значимый параллелизм на уровне одного модуля памяти (за счёт распределения срезов весов на несколько вычислительных трактов и дублирования по активациям).

В обычном режиме СФ-блок на базе двухпортовой блочной ОЗУ функционирует как стандартная память. Переход к выполнению вычислений не задаётся статическим битом конфигурации, а инициируется записью в заранее зарезервированный командный диапазон адресов на одном из портов. Встроенный конечный автомат (ВКА) постоянно контролирует обращения к этому диапазону: если адрес попадает в него, транзакция трактуется как команда вычислений в памяти, в противном случае – как обычная операция чтения/записи.

Пока командные записи не поступают, ВКА остаётся неактивным, и поведение пользовательского

порта полностью соответствует стандартной памяти; семантика штатных сигналов не изменяется. Команда (ВП-инструкция) передаётся в виде одной или нескольких записей по адресам из командного диапазона, причём поле данных инструкции содержит параметры операции: тип/режим выполнения, адреса и размеры областей с весовыми коэффициентами и активациями, настройки точности, флаги параллелизма/дублирования.

Получив и декодировав такую инструкцию, ВКА формирует требуемую последовательность управляющих сигналов для встроенной вычислительной логики и интерфейса памяти, организуя чтение весов/активаций, выполнение УНК-операций и запись результатов. Доступ к остальной адресной области вне командного диапазона продолжается в штатном режиме, включая параллельную работу второго порта памяти.

Архитектура вычислительных ядер внутри блочной ОЗУ

Основу вычислительного ядра внутри блочной ОЗУ составляет массив памяти и ассоциированная с ним периферийная программируемая логика. Структурная схема разработанного вычислительного ядра представлена на рисунке 3.

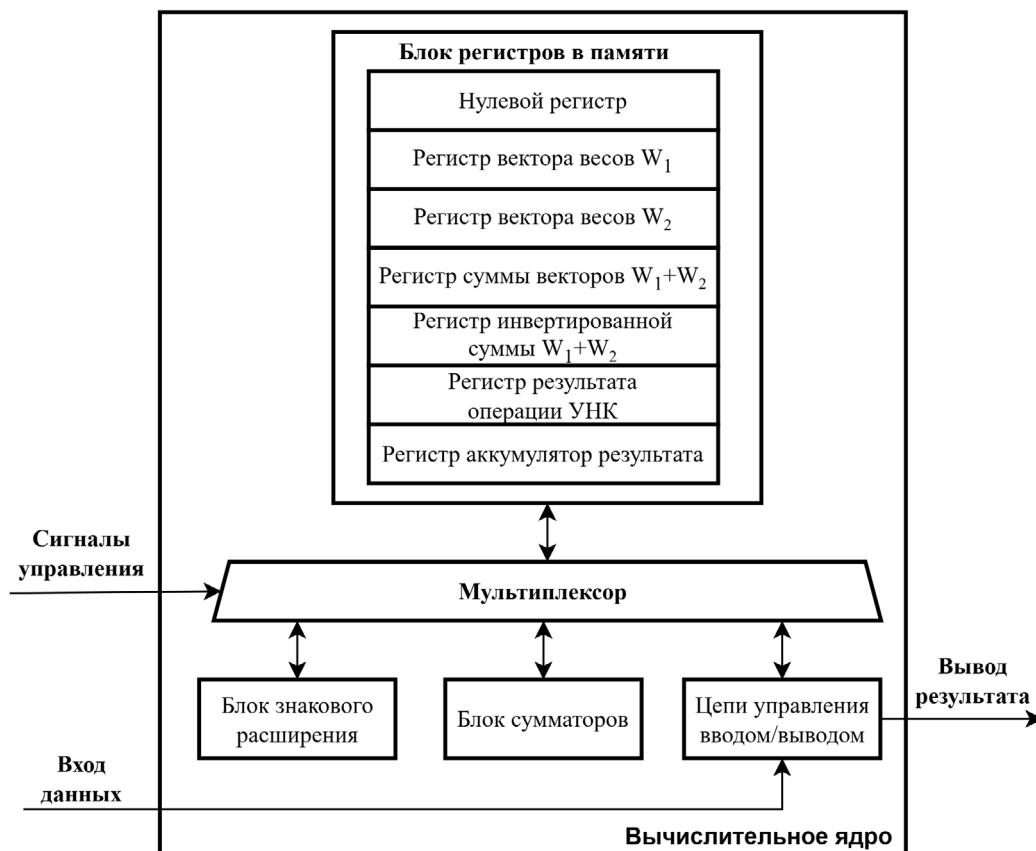


Рис. 3. Структурная схема вычислительного ядра

Регистровый массив вычислительного ядра представляет собой набор регистров, организованный в виде структуры, имеющей фиксированную 7-строчную организацию и конфигурируемое количество столбцов-регистров. Каждый регистр имеет предопределеное функциональное назначение:

1. **Нулевой регистр.** Аппаратно сконфигурирован для хранения нулевых значений, используемых в качестве одного из операндов при определенных комбинациях битов активаций.

2. **Регистры вектора весов W_1 и W_2 .** Предназначены для размещения векторов весов W_1 и W_2 , предварительно скопированных из основного массива блочных ОЗУ.

3. **Регистр ($W_1 + W_2$).** Содержит вычисленную по-элементную сумму векторов W_1 и W_2 .

4. **Регистр инверсной суммы.** Используется для хранения в своих регистрах инвертированных значений частичных сумм, необходимых для корректной реализации операции вычитания при обработке старшего значащего бита операндов в представлении с дополнением до двух.

5. **Строка регистров результата двойной операции с УНК.** Аккумулирует в своих регистрах результат выполнения поэлементных двух совмещенных операций УНК вида: $P_j = (W_{1j} \cdot I_1 + W_{2j} \cdot I_2)$, где:

– P_j – j-й элемент вектора результата двойной операции с УНК;

– W_{1j} , W_{2j} – j-е элементы векторов весов W_1 и W_2 ;

– I_1 , I_2 – скалярные значения текущих битов активаций.

6. **Регистр аккумулятор** предназначен для по-элементного суммирования результатов нескольких последовательных двойных операций с УНК.

Помимо регистрового массива, каждое вычислительное ядро внутри блочной ОЗУ включает в себя набор специализированных периферийных логических схем. Эти схемы обеспечивают поддержку операций над данными, хранящимися в регистрах, управление потоками данных и синхронизацию работы ядра с остальными компонентами архитектуры. Ниже приведены ключевые периферийные схемы.

1. **Блок знакового расширения** осуществляет знаковое расширение векторов весов, считанных из основного массива блочных ОЗУ, до разрядности, необходимой для корректной работы сумматора с одиночным потоком команд и множественным потоком данных и предотвращения переполнения в процессе вычислений, перед их записью в соответствующие строки регистров W_1 и W_2 .

2. **Набор параллельно работающих полных сумматоров**, разрядность и количество которых

соответствует числу столбцов-регистров в регистровом массиве ядра. Он необходим для выполнения битово-параллельного сложения множества элементов векторов весов (считанных из строк регистров весов или строк нулевого регистра) с текущим битом активации (или их комбинацией) и последующего накопления результатов в регистре результата. Конфигурируемость сумматора позволяет эффективно оперировать с данными различной точности, хранящимися в регистрах.

3. **Схемы управления чтением/записью регистра** **строго** **массива ядра** управляют сигналами разрешения записи и выбора для каждого регистра, а также операциями записи результатов вычислений обратно в регистры.

4. **Мультиплексоры** обеспечивают коммутацию потоков данных между основным массивом блочных ОЗУ, регистровым массивом ядра и сумматорами.

Задержка выполнения одной двойной операции с УНК в вычислительном ядре линейно зависит от разрядности входных активаций и составляет $(n+2)$ такта системной частоты для синхронного варианта ядра, либо $(n/2+2)$ такта для варианта с работой на удвоенной тактовой частоте ядра.

Таким образом, вычислительные ядра внутри блочной ОЗУ в рассматриваемой архитектуре вычислений в памяти представляют собой эффективный вычислительный конвейер, обеспечивающий высокую степень параллелизма вычислений и гибкости при интеграции в аппаратные ускорители вывода ГНС на базе ПЛИС.

АППАРАТНЫЙ УСКОРИТЕЛЬ ТРАНСФОРМЕРНОЙ НЕЙРОННОЙ СЕТИ VISION TRANSFORMER

Для демонстрации применения предложенной архитектуры СФ-блоков рассмотрено проектирование аппаратного ускорителя модуля самовнедления трансформерной модели Vision Transformer в базовой конфигурации (ViT-Base). Этот пример иллюстрирует возможности гибридной вычислительной архитектуры при обработке современных трансформерных моделей, активно применяемых в задачах компьютерного зрения.

В базовой конфигурации Vision Transformer (ViT-Base) входное изображение размером 224x224 пикселя разбивается на непересекающиеся патчи размером 16x16 пикселей, что формирует последовательность из $14 \times 14 = 196$ патчей. С добавлением служебного токена классификации общая длина последовательности составляет 197 элементов.

Каждый патч проецируется в пространство признаков размерностью 768 (размерность эмбеддинга $d_{model}=768$). Механизм многоголового самовнимания (multi-head self-attention) использует 12 голов ($h=12$), каждая из которых оперирует в пространстве размерностью $d_k=d_v=d_{model}/h=64$. Эти параметры являются стандартными для модели ViT-Base и широко используются в открытых реализациях [20].

Вычислительные операции в модуле самовнимания

Модуль самовнимания (self-attention) выполняет следующую последовательность операций. На первом этапе производятся линейные проекции входной последовательности для формирования матриц запросов (**Q**), ключей (**K**) и значений (**V**). Эта операция представляет собой матрично-матричное умножение размерности $[197 \times 768] \times [768 \times 2304]$, где выходная размерность 2304 соответствует конкатенации трёх проекций. После получения матриц **Q**, **K** и **V** они разделяются по головам внимания.

Внутри каждого модуля самовнимания выполняются следующие операции. Вычисление матрицы оценок внимания осуществляется умножением запросов на транспонированные ключи: $[197 \times 64] \times [64 \times 197]$, результатом чего является матрица оценок размерности $[197 \times 197]$. После применения операции softmax к каждой строке матрицы оценок (что может быть реализовано на блоках цифровой обработки сигналов или специализированной логике) полученные нормализованные веса внимания умножаются на матрицу значений $[197 \times 197] \times [197 \times 64]$, формируя выходную матрицу модуля размерностью $[197 \times 64]$. Результаты от всех 12 голов конкатенируются, образуя тензор размерности $[197 \times 768]$, который затем проходит через конечную линейную проекцию: $[197 \times 768] \times [768 \times 768]$.

Распределение вычислительной нагрузки между СФ-блоками и блоками цифровой обработки сигналов

Предлагаемая архитектура позволяет эффективно распределить вычислительную нагрузку модуля самовнимания между СФ-блоками с вычислительными ядрами и традиционными блоками цифровой обработки сигналов. Линейные проекции для формирования **Q**, **K**, **V**, а также конечная проекция могут выполняться на СФ-блоках с использованием квантованных весовых коэффициентов разрядностью 2/4/8 бит и активаций разрядностью 2–8 бит в зависимости от требований к точности для конкретного слоя. Операция умножения **Q** \times **K** \wedge **T** и последующее умножение на **V** также могут быть эф-

ективно реализованы на СФ-блоках, особенно при использовании квантования с переменной разрядностью, где различные слои или компоненты могут иметь различную разрядность.

Операции, требующие повышенной точности, такие как вычисление softmax, нормализация слоёв и, при необходимости, первичные и выходные проекции в повышенной разрядности, выполняются на блоках цифровой обработки сигналов. Это распределение обусловлено тем, что СФ-блоки наиболее эффективны для массовых операций умножения-накопления с низкоразрядными данными, в то время как блоки цифровой обработки сигналов обеспечивают гибкость для операций с произвольной точностью и специализированных функций. Принцип распределения разрядности по слоям определяется на этапе квантования модели с учётом чувствительности различных компонентов самовнимания к снижению разрядности и может быть оптимизирована для конкретных задач классификации изображений.

Организация потоков данных и конвейерная буферизация

Ключевой особенностью предлагаемой архитектуры СФ-блоков является сохранение одного порта двухпортовой памяти доступным для стандартных операций чтения и записи во время выполнения вычислений в памяти. Это позволяет реализовать двойную буферизацию без необходимости использования внешних буферных структур. Во время выполнения операций умножения-накопления на вычислительных ядрах СФ-блока через порт А другие компоненты системы, включая блоки цифровой обработки сигналов и управляющую логику, могут обращаться к данным через порт Б. Это обеспечивает конвейеризацию операций загрузки данных, вычислений и выгрузки результатов, что критично для достижения высокой эффективности использования аппаратуры и минимизации простоеев вычислительных ресурсов.

Организация параллельной работы СФ-блоков и блоков ЦОС при обработке последовательности тайлов показана на рис. 4. Три стадии конвейера (загрузка данных следующего тайла, вычисление текущего тайла и выгрузка результатов предыдущего тайла) выполняются одновременно, что обеспечивает высокую эффективность использования аппаратных ресурсов.

Поскольку размерность матричных операций в модуле самовнимания ViT-Base (например, $[197 \times 768] \times [768 \times 2304]$ для линейных проекций **Q/K/V**) существенно превышает объём доступной

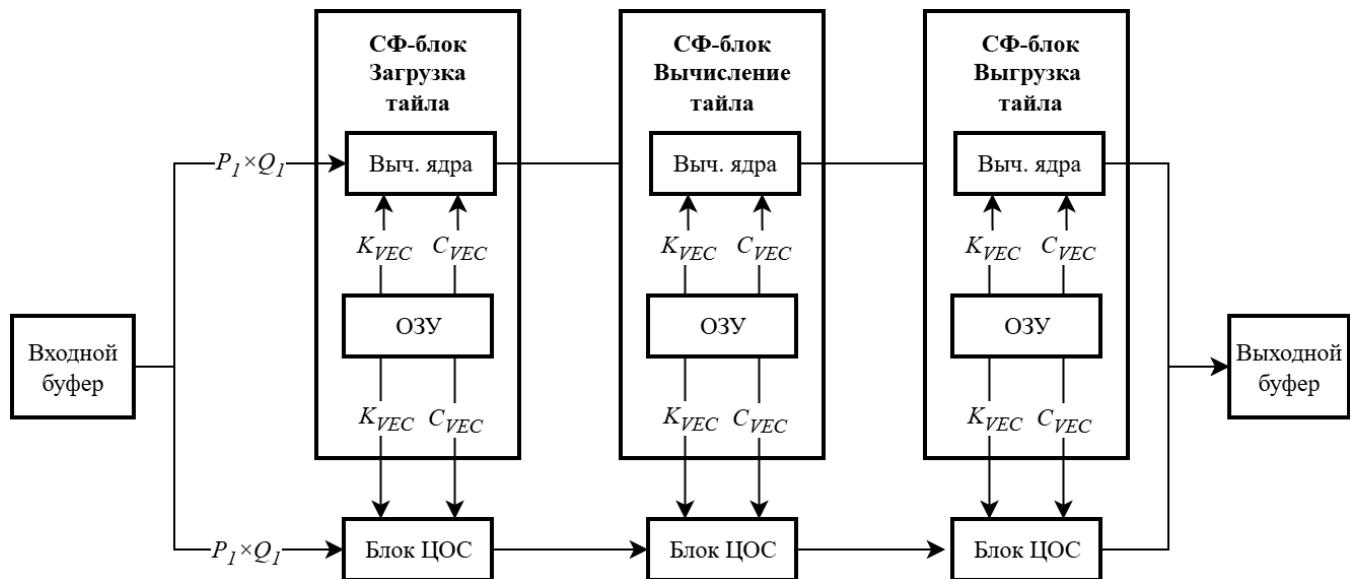


Рис. 4. Конвейерная организация гибридного ускорителя с двойной буферизацией для обработки последовательности тайлов в модуле самовнимания *Vision Transformer*

блочной памяти и число вычислительных ресурсов на кристалле ПЛИС, выполнение операций организуется путём декомпозиции на тайлы — фрагменты данных фиксированного размера, последовательно обрабатываемые доступными СФ-блоками и блоками ЦОС. Размеры тайлов определяются параметрами K_{vec} (число выходных каналов), C_{vec} (число входных каналов) и другими размерностями, оптимизируемыми для максимизации загрузки аппаратуры при заданных ограничениях на объём памяти и число вычислительных блоков. Такая организация потоков данных позволяет параллельно выполнять несколько операций: загрузку весовых коэффициентов и активаций следующего тайла через порт записи СФ-блока; выполнение операций умножения-накопления на вычислительных ядрах текущего тайла; чтение результатов предыдущего тайла и их объединения на блоках цифровой обработки сигналов через порт чтения СФ-блока. Встроенный конечный автомат СФ-блока управляет синхронизацией этих операций и арбитражем доступа к портам памяти.

Аппаратные ограничения и оценка производительности

Производительность аппаратного ускорителя ViT-Base на базе предлагаемой архитектуры ограничивается несколькими факторами. Основным ограничивающим ресурсом является количество доступных примитивов блочной оперативной памяти (например, RAMB36E2 в семействе Xilinx UltraScale+), которые могут быть модифицированы в СФ-блоки. Каждый СФ-блок с четырьмя вычисли-

тельными ядрами размером 7×32 столбца занимает площадь, превышающую базовый примитив RAMB36E2 на 19,6 % для варианта с ядрами 7×64 столбца. Для платформы Xilinx Virtex UltraScale+ VU9P, содержащей 2160 примитивов RAMB36, теоретически может быть реализовано аналогичное количество СФ-блоков, обеспечивающих высокую вычислительную плотность.

Частота тактирования системы также является критическим параметром. Моделирование в инструменте COFFEE с использованием 22-нанометровой предиктивной технологической модели показывает, что наибольшая задержка распространения в вычислительном ядре СФ-блока-М составляет 903 пс, что соответствует максимальной частоте около 1,1 ГГц для изолированного блока.

Однако при двойном тактировании вычислительных ядер (работа на удвоенной частоте относительно главного массива блочной оперативной памяти) максимальная частота главного массива ограничивается до 553 МГц, что соответствует снижению на 24% относительно базовой частоты примитива RAMB36E2. В реалистичных условиях синтеза на программируемой логической интегральной схеме частота системы дополнительно ограничивается задержками маршрутизации и программируемой логики, и типичные ускорители глубоких нейронных сетей на программируемых логических интегральных схемах семейства UltraScale+ работают на частотах 200–400 МГц.

Пропускная способность внутренних шин данных между СФ-блоками, блоками цифровой обработки сигналов и буферами также влияет на об-

щую производительность системы. При ширине шины данных 32 бита (стандартная конфигурация для блочной оперативной памяти с глубиной 512 и шириной 32 бита) и рабочей частоте 400 МГц пропускная способность одного порта составляет 12,8 Гбайт/с. Для модуля самовнимания ViT-Base с 197 токенами и эмбеддингом 768 объём данных на один проход составляет приблизительно 151 Кбайт для матрицы входных активаций (при 8-битном квантовании), что позволяет загрузить данные за 11,8 мкс при указанной пропускной способности.

Объём обмена между памятью и вычислительной логикой значительно снижается за счёт выполнения операций умножения-накопления непосредственно в СФ-блоках. Традиционная архитектура требует для каждой операции умножения-накопления чтения весового коэффициента и активации из памяти, передачи их в удалённый блок цифровой обработки сигналов и записи результата обратно.

В предлагаемой архитектуре весовые коэффициенты остаются в СФ-блоке, и только активации и конечные результаты передаются через интерфейсы памяти. Для операции матрично-матричного умножения размерности $[M \times K] \times [K \times N]$ традиционный подход требует $(M \times K \times N)$ операций чтения весов и активаций, в то время как предлагаемая архитектура требует только $(M \times K)$ операций чтения активаций и $(K \times N)$ операций записи результатов, что обеспечивает снижение объёма обмена пропорционально параметру N .

Оценка производительности для ViT-Base с использованием 6-битных активаций и 8-битных весовых коэффициентов (конфигурация, обеспечивающая менее 0,5 % потери точности по сравнению с моделью с плавающей точкой на задаче классификации ImageNet, как показано в разделе 4) проводилась с использованием моделирования с точностью до такта. При использовании 1000 СФ-блоков с двойным тактированием вычислительных ядер на частоте 400 МГц главного массива и 800 МГц вычислительных ядер пиковая пропускная способность для 6-битных операций умножения-накопления составляет приблизительно 25,6 ТОП/с (тераопераций в секунду). С учётом 648 блоков цифровой обработки сигналов (конфигурация, аналогичная Intel Stratix-10 GX400), каждый из которых выполняет две 8-битные операции умножения-накопления за такт при уплотнении, дополнительная пропускная способность блоков цифровой обработки сигналов составляет 0,52 ТОП/с. Суммарная пиковая пропускная способность гибридной системы достигает приблизительно 26,1 ТОП/с для конфигурации с переменной разрядностью. Для 2-битных операций

умножения-накопления при использовании всех 2160 доступных СФ-блоков пиковая пропускная способность может достигать 50,7 ТОП/с.

Время выполнения одного прохода модуля самовнимания ViT-Base оценивается с учётом декомпозиции операций на тайлы и конвейеризации. Для линейной проекции $Q/K/V$ размерности $[197 \times 768] \times [768 \times 2304]$ при разбиении на тайлы с размером выходного канала $K_{vec} = 64$ и размером входного канала $C_{vec} = 96$ (выбранные для пропорциональной загрузки СФ-блоков и блоков цифровой обработки сигналов) требуется выполнение 288 тайловых операций. Каждая тайловая операция размерности $[197 \times 96] \times [96 \times 64]$ включает 1,208,832 операций умножения-накопления. При задержке 5 тактов на одну операцию MAC2 (двойное умножение-накопление) в режиме двойного тактирования и частоте вычислительных ядер 800 МГц время выполнения одной тайловой операции составляет приблизительно 3,78 мкс. С учётом конвейеризации и частичного перекрытия загрузки, вычислений и выгрузки общее время выполнения проекции $Q/K/V$ оценивается в диапазоне 1,1–1,5 мс для одного изображения.

Аналогичные оценки для операций умножения $Q \times K \times T$ и применения весов внимания к V дают суммарное время вычисления модуля самовнимания на уровне 3–5 мс на одно изображение при использовании конфигурации с переменной разрядностью, что соответствует пропускной способности 200–330 кадров в секунду для изолированного модуля самовнимания. Полная модель ViT-Base содержит 12 последовательных блоков трансформера, каждый из которых включает модуль самовнимания и слой прямого распространения (feed-forward), что масштабирует общее время выполнения пропорционально. С учётом дополнительных операций нормализации и обходных (residual) соединений пропускная способность для полной модели ViT-Base оценивается в диапазоне 20–30 кадров в секунду на рассматриваемой конфигурации программируемой логической интегральной схемы, что является конкурентоспособным показателем для встраиваемых систем компьютерного зрения.

Методика оценки энергопотребления требует дополнительной экспериментальной верификации на физической платформе и является предметом дальнейших исследований. Предварительные оценки на основе моделей энергопотребления блочной оперативной памяти и вычислительных ядер указывают на потенциальное снижение энергозатрат на операцию умножения-накопления на 30–50 % по сравнению с традиционной архитек-

турой с раздельными блоками памяти и блоками цифровой обработки сигналов за счёт сокращения объёма передачи данных.

РЕЗУЛЬТАТЫ

Для оценки предложенной архитектуры и ее сопоставления с существующими решениями было применено моделирование для анализа производительности на уровне регистровых передач для ПЛИС и конкретных приложений.

Для моделирования занимаемой площади кристалла и задержек сигналов в блочной ОЗУ разработанного СФ-блока, включая базовые компоненты и новые функциональные блоки, использовался инструмент расчета COFFE. Моделирование в COFFE проводилось с использованием 22-нм предиктивной технологической модели (PTM), такой режим принят как репрезентативный для современных узлов, близких к 16-нм классу устройств семейства UltraScale+ [21].

Оценка пиковой производительности и издержек на уровне всей ПЛИС производилась на основе эталонных архитектур ПЛИС, таких как Xilinx Virtex UltraScale+, с использованием данных о количестве и относительной площади различных ресурсов (конфигурируемых логических блоков (CLB)), бло-

ков ЦОС - DSP, блоков ОЗУ - Block RAM, например, RAMB36), полученных из открытых источников, документации Xilinx (AMD). Пиковая пропускная способность операций умножения и накопления (УНК) для программируемой логики (реализованной на CLB) и стандартных блоков ЦОС определялась путем синтеза в Xilinx Vivado Design Suite и последующего экстраполирования на все доступные ресурсы. Для предыдущих архитектур пиковая производительность рассчитывалась на основе количества параллельно выполняемых операций, их задержки и максимальной рабочей частоты модифицированных блоков ОЗУ, полученной из моделирования в COFFE (Circuit Optimization for FPGA Exploration — инструмент оптимизации схем для исследования ПЛИС) или заявленной в работах-источниках [17, 18].

Далее приведена сводная таблица, аккумулирующая ключевые характеристики и показатели производительности для рассматриваемых архитектур, включая eDSP (модифицированный блок ЦОС с расширенными возможностями), CCB (Compute Cache Block — вычислительный кэш-блок, ранняя архитектура с вычислительными ядрами в памяти) и архитектура блочной оперативной памяти с встроенными вычислительными ядрами и поддержкой обработки данных разнородной точности.

Таблица 1

Сравнительные характеристики архитектур аппаратных ускорителей на базе модифицированных блоков ПЛИС

| Характеристика | eDSP [17] | CCB [18] | Данная статья |
|---|--------------|----------------------|---------------|
| Модифицируемый блок ПЛИС | Блок ЦОС | Блок ОЗУ | Блок ОЗУ |
| Поддерживаемая точность УНК (бит) | 4, 8 | Произвольная | 2,4,8 |
| Доп. затраты площади (на блок) | 12% | 16.80% | 19.60% |
| Доп. затраты площади (на ядро ПЛИС) | 1.10% | 3.40% | ~5.7% |
| Снижение макс. частоты блока ОЗУ | 0% (для ЦОС) | 60% | ~32% |
| Количество MAC на операцию / Задержка (циклы) | | | |
| 2-бита | 8/1 | 160 / 16 (беззнак.) | 64 / 3 |
| 4-бита | 8/1 | 160 / 42 (беззнак.) | 32 / 4 |
| 8-бит | 8/1 | 160 / 113 (беззнак.) | 16 / 6 |
| Пиковая пропускная способность УНК (TOP/c) всей ПЛИС | | | |
| 2-бита | ~5.2 | ~19 | ~50.7 |
| 4-бита | ~3.7 | ~10 | ~19.8 |
| 8-бит | ~1.8 | ~4 | ~6.9 |

Ключевым достижением предложенной архитектуры является значительно более высокая пиковая пропускная способность операций умножения-накопления по сравнению с eDSP и ССВ во всех рассмотренных точностях и особенно для операций с низкой точностью (2- и 4-битные веса/активации).

ЗАКЛЮЧЕНИЕ

В данной работе представлена и оценена новая архитектура аппаратного ускорителя ГНС на ПЛИС, нацеленная на решение проблем неэффективного использования ресурсов стандартных блоков цифровой обработки сигналов при работе с низкоразрядными данными и преодоление ограничений «стены памяти».

Основой предложенного решения являются СФ-блоки на базе блочной ОЗУ, позволяющие выполнять часть операций непосредственно у данных и, таким образом, снижать объём и частоту передач между памятью и логикой, при этом обеспечивая эффективную обработку данных с переменной раз-

рядностью (весовые коэффициенты 2/4/8 бит; активации 2–8 бит).

Сравнительный анализ с известными архитектурами, такими как eDSP и ССВ, продемонстрировал значительное превосходство разработанного подхода в пиковой производительности, особенно для операций с 2- и 4-битной точностью, где прирост достигает нескольких раз (до 50,7 ТОП/с для 2-битных операций).

Полученные результаты открывают перспективы для создания высокопроизводительных и энергоэффективных ПЛИС-ускорителей для сложных глубоких нейронных сетей, позволяя применять более глубокое снижение разрядности.

Дальнейшие направления исследований включают оптимизацию управляющей логики, перенос в архитектуру системы-на-кристалле на базе ПЛИС, разработку средств автоматизированного проектирования и отображения ГНС на предложенную архитектуру, а также экспериментальную оценку энергопотребления.

СПИСОК ЛИТЕРАТУРЫ

1. Strielkowski W., Vlasov A., Selivanov K., Muraviev K., Shakhnov V. Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: a review // Energies. 2023. Т. 16, № 10. Ст. 4025.
2. Prudius A. A., Karpunin A. A., Vlasov A. I. Analysis of machine learning methods to improve efficiency of big data processing in industry 4.0 // Journal of Physics: Conference Series. 2019. Т. 1333. Ст. 032065.
3. Lammie C., Olsen A., Carrick T., Rahimi Azghadi M. Low-Power and High-Speed Deep FPGA Inference Engines for Weed Classification at the Edge // IEEE Access. 2019. Т. 7. С. 51171–51184. DOI: 10.1109/ACCESS.2019.2911709.
4. Власов А. И., Иванов В. В., Косолапов И. А. Методы упреждающего прогнозирования состояния широкополосной сети связи // Программные продукты и системы. 2011. № 1. С. 3–6.
5. Власов А. И. Аппаратная реализация нейровычислительных управляющих систем // Приборы и системы. Управление, контроль, диагностика. 1999. № 2. С. 61–65.
6. Жалялов Р. Р. Использование свёрточных нейронных сетей для решения задач распознавания образов в промышленности // Информационные технологии в проектировании и производстве. 2023. № 2 (190). С. 21–31.
7. Kolli H., Rao S. J., Pandian V., Suganthi R., Rama R. S. Optimizing Neural Network Inference in Edge Robotics by Harnessing FPGA Hardware Acceleration // Journal of Electrical Systems. 2024. Т. 20, № 5s. С. 1935–1943. DOI: 10.52783/jes.2529.
8. Пирогова М. А., Лешихина И. Е., Шведов Е. И., Краюшкин В. А. Практика использования компьютерного зрения и искусственного интеллекта для автоматизации процессов визуального контроля качества выполнения сборочных операций // Информационные технологии в проектировании и производстве. 2023. № 4 (192). С. 11–19.
9. Wan Z., Lele A., Yu B., Liu S., Wang Y., Reddi V. J., Hao C., Raychowdhury A. Robotic Computing on FPGAs: Current Progress, Research Challenges, and Opportunities // 2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS). 2022. С. 291–295. DOI: 10.1109/AICAS54282.2022.9869951.
10. Шахнов В. А., Власов А. И., Поляков Ю. А., Кузнецов А. С. Нейрокомпьютеры: архитектура и схемотехника. М.: Машиностроение, 2000. 64 с. (Серия 9. Приложение к журналу «Информационные технологии».)
11. Mittal S. A Survey of FPGA-based Accelerators for Convolutional Neural Networks // Neural Computing and Applications. 2020. Т. 32, вып. 4. С. 1189–1210. DOI: 10.1007/s00521-018-3761-1.

- 12.**Le Blevec H., Léonardon M., Tessier H., Arzel M. Pipelined Architecture for a Semantic Segmentation Neural Network on FPGA // 2023 30th IEEE International Conference on Electronics, Circuits and Systems (ICECS). 2023. C. 1–4. DOI: 10.1109/ICECS58634.2023.10382715.
- 13.**Zhang J., Cheng L., Li C., Li Y., He G., Xu N., Lian Y. A Low-Latency FPGA Implementation for Real-Time Object Detection // 2021 IEEE International Symposium on Circuits and Systems (ISCAS). 2021. C. 1–5. DOI: 10.1109/ISCAS51556.2021.9401577.
- 14.**Yu J., Xu Z., Zeng S., Yu C., Qiu J., Shen C., Xu Y., Dai G., Wang Y., Yang H. INCA: INterruptible CNN Accelerator for Multi-tasking in Embedded Robots // 57th ACM/IEEE Design Automation Conference (DAC). 2020. C. 1–6. DOI: 10.1109/DAC18072.2020.9218717.
- 15.**Wulf W. A., McKee S. A. Hitting the Memory Wall: Implications of the Obvious // ACM SIGARCH Computer Architecture News. 1995. Т. 23, № 1. С. 20–24.
- 16.**Власов А. И., Кузьмина С. В., Макаров С. С., Петровичев Ф. М., Шахнов В. А. Анализ решений двойного применения вычислительных средств на платформе «Эльбрус» // Нейрокомпьютеры: разработка, применение. 2023. Т. 25, № 1. С. 46–62.
- 17.**Rasoulinezhad S., Zhou H., Wang L., Leong P. H. W. PIR-DSP: An FPGA DSP Block Architecture for Multi-precision Deep Neural Networks // Proc. IEEE 27th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). San Diego, CA, USA, 2019. C. 35–44. DOI: 10.1109/FCCM.2019.00015.
- 18.**Wang X., Goyal V., Yu J., Bertacco V., Boutros A., Nurvitadhi E., Augustine C., Iyer R. R., Das R. Compute-Capable Block RAMs for Efficient Deep Learning Acceleration on FPGAs // Proc. IEEE 29th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). 2021. C. 88–96.
- 19.**Аверьянишин А. Е., Власов А. И., Евдокимова Е. В. Иерархическая пирамидальная субдискретизация в глубоких свёрточных сетях для распознавания визуальных образов // Нейрокомпьютеры: разработка, применение. 2021. Т. 23, № 1. С. 17–31.
- 20.**Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // International Conference on Learning Representations (ICLR). 2021. URL: <https://openreview.net/forum?id=YicbFdNTTy> (дата обращения: 19.10.2025).
- 21.**AMD (Xilinx). UltraScale+ FPGAs. Data Sheet: Overview (DS925) [Электронный ресурс]. URL: <https://docs.amd.com/r/en-US/ds925-ultrascale-plus-overview> (дата обращения: 11.10.2025).

УДК: 004.42, 004.75

Программные решения для анализа данных в правоохранительной деятельности

V.L. Akapurov, A.V. Borisenko

Software Solutions for Data Analysis in Law Enforcement

Abstract. The article is devoted to the study of the state of information and analytical support for law enforcement agencies in the context of digitalization and explores the most effective way to address information analytics issues in law enforcement. This article analyzes the accumulated experience of using disparate automated information systems, confirms the feasibility of using integrated software, formulates the required set of software platform functions, and justifies the choice of a specific software solution. Based on this research, it is concluded that there is a pressing need to improve the effectiveness of predictive analytics in law enforcement agencies using integrated software platforms. DataWalk software is proposed as such a solution.

Keywords: law enforcement, data analysis, software, automated information systems, software platforms.

Аннотация. Статья посвящена исследованию состояния информационно-аналитического обеспечения правоохранительных органов в условиях цифровизации общества и обоснованию наиболее эффективного способа решения проблем информационной аналитики в правоохранительной деятельности. Проведен анализ накопленного опыта использования разрозненных автоматизированных информационных систем, подтверждена целесообразность применения интегрированного программного обеспечения, сформулирован требуемый набор функций программной платформы и обоснован выбор конкретного программного решения. На основании проведенного исследования сделан вывод об актуальной необходимости повышения эффективности прогнозной аналитики правоохранительных органов на базе интегрированных программных платформ. В качестве такого решения предлагается использовать программное обеспечение DataWalk.

Ключевые слова: правоохранительная деятельность, анализ данных, программное обеспечение, автоматизированные информационные системы, программные платформы.

ВВЕДЕНИЕ

Каждое структурное подразделение МВД России может извлечь необходимые сведения для повышения эффективности правоохранительной деятельности из анализа криминогенной обстановки и природы противоправных действий. Патрульные получают представление о закономерностях совершения преступлений в своём районе или городе, что помогает им предвидеть инциденты. Оперативные сотрудники используют анализ для выявления тенденций совершения преступлений, имеющих решающее значение для организации проведения оперативно-розыскных мероприятий и их расследования. Руководящий состав использует его для более эффективного распределения ресурсов и стратегического планирования, после чего может

направить больше сотрудников и ресурсов в районы и зоны повышенного риска.

Анализ данных о преступности – важнейшая функция правоохранительных органов, в рамках которой аналитики используют данные, статистические методы и картографирование преступности для понимания и прогнозирования преступной деятельности. Такой подход, основанный на данных, помогает правоохранительным органам оценивать меры реагирования и совершенствовать стратегии профилактики. Однако реализация информационно-аналитического обеспечения правоохранительной деятельности сталкивается с рядом проблем, решение которых можно возложить на специализированное программное обеспечение, при этом особого внимания заслуживают интеллектуальные программные платформы для анализа преступности.

В настоящей работе исследуется состояние информационно-аналитического обеспечения правоохранительных органов в условиях цифровизации общества с целью обосновать наиболее эффективный способ решения проблем информационной аналитики в правоохранительной деятельности. В рамках поставленной цели было необходимо решить следующие задачи: проанализировать опыт использования разрозненных автоматизированных информационных систем локального, регионального и федерального уровней; обосновать необходимость использования интегрированного программного обеспечения; сформулировать требуемый функционал программной платформы; обосновать выбор конкретного программного решения.

ПРОБЛЕМЫ И ИНСТРУМЕНТЫ ИНФОРМАЦИОННОЙ АНАЛИТИКИ

В настоящее время метод «ручного» анализа огромного количества доступных данных в правоохранительных органах непрактичен и неэффективен и сталкивается с рядом проблем:

1. **Различные форматы данных.** Данные из таких источников, как статистические отчёты органов внутренних дел, социальные сети и системы видеонаблюдения, поступают в разных форматах, включая структурированные и неструктурированные данные, которые традиционные решения зачастую не поддерживают.

2. **Большой объём данных.** С появлением цифровых технологий объём данных, с которыми приходится работать правоохранительным органам, увеличился в геометрической прогрессии. Огромный объём данных может быть ошеломляющим и приводить к тому, что важная информация остаётся незамеченной.

3. **Разрозненность данных,** когда информация хранится в отдельных системах и банках данных даже в рамках одной организации, а также в разных правоохранительных органах, что может препятствовать получению полного представления о тенденциях и закономерностях преступности.

Использование значительного количества разрозненных автоматизированных информационных систем различного ведомственного подчинения характеризуется низкой эффективностью и невозможностью реализации доктрины формирования единого информационного правоохранительного пространства, поэтому аксиомой настоящего этапа развития правоохранительной деятельности в условиях цифровизации общества является не просто

необходимость использования информационных технологий для реализации, в частности, информационно-аналитического обеспечения, как ключевого направления оценки эффективности борьбы с преступностью, но обязательный переход на применение универсальных программных платформ.

Данный вывод следует из анализа использования программного обеспечения (ПО) правоохранительными органами как в России, так и за рубежом. Согласно недавнему исследованию Cognyte, посвящённому аналитике данных для правоохранительных органов, большинство правоохранительных органов по-прежнему используют несколько разрозненных решений для анализа данных [1].

Ограничения и неэффективность такого подхода связаны с необходимостью сотрудников правоохранительных органов повседневно разбираться в лабиринте цифровых инструментов, требуемых для сбора и анализа оперативной информации. То, что должно упростить и увеличить эффективность работы оперативных сотрудников, превращается в трудоемкое взаимодействие с различными системами, усугубляемое недостаточным уровнем их информационно-технологической компетентности [2].

Это идеальный пример «синдрома вращающегося кресла», когда переключение между несколькими системами превращается в жонглирование, которое не только отнимает время, но и превращает потенциальные преимущества в громоздкие препятствия.

Только из открытых источников можно выделить целый ряд автоматизированных информационных систем (АИС) локального, регионального и федерального уровней, применяемых в органах внутренних дел (ОВД) МВД России и других правоохранительных структурах Российской Федерации (табл. 1).

Указанные системы разрабатывались на рубеже веков с использованием средства создания баз данных и работы с ними FLINT, что отразилось на крайне недружественном, предъявляющем довольно жесткие требования к базовой подготовке пользователей интерфейсе, частых сбоях и возможном негативном воздействии вредоносного программного обеспечения. Для примера достаточно сослаться на клиентские требования АИПС «Аэропорт»: программные приложения для Windows 32 bit, совместимым с ОС Windows XP + FrameWork 2.0, Windows 7 и выше.

Таблица 1

АИС правоохранительных органов

| Система | Назначение | Возможности |
|---|--|--|
| Автоматизированная система информационного обеспечения (АСИО) «Прокуратура» | Создание на всей территории Российской Федерации единой автоматизированной информационной среды. | Организация автоматизированных рабочих мест по каждому виду деятельности органов прокуратуры, формирование локальных проблемно-ориентированных баз данных и программно-технических средств для решения определенной логически полной группы задач по каждому направлению деятельности. |
| Автоматизированная информационно-поисковая система (АИПС) «Автотоиск» | Получение информации об угнанных, задержанных, похищенных средствах автотранспорта. | Отбор и вывод информации по заданному в запросе условию. |
| АИПС «Аэропорт» | Выявление в аэропортах на основе паспортных данных лиц, представляющих оперативный интерес, а также утраченных паспортов. | Обработка данных в реальном времени. |
| АИПС «Криминал-И» | Учёт правонарушений, совершённых иностранцами и лицами без гражданства. | Отбор и вывод информации по заданному в запросе условию. |
| АИПС «Опознание» | Получение информации о лицах, пропавших без вести, о неопознанных трупах, о неизвестных больных и детях. | Отбор и вывод информации по заданному в запросе условию. |
| АИПС «ФР-Оповещение» | Учет преступников, разыскиваемых поискам предприятий и организаций (госдолжников), граждан – неплательщиков алиментов и граждан, пропавших без вести. Отработка запросов на лиц, находящихся в федеральном розыске, подготовка циркуляров на объявление или прекращение розыска. | Отбор и вывод информации по заданному в запросе условию. |
| АИПС «Антиквариат» | Предоставление сведений об утраченных и выявленных предметах, представляющих историческую, художественную или научную ценность. | Отбор и вывод информации по заданному в запросе условию. |
| АИПС «Вещь» | Предоставление сведений о похищенных и изъятых номерных вещах, документах, ценных бумагах общегосударственного обращения. | Отбор и вывод информации по заданному в запросе условию. |
| АИПС «Сейф» | Сбор, обработка и выдача сведений о преступлениях, при совершении которых взламывались металлические хранилища. | Отбор и вывод информации по заданному в запросе условию. |
| Автоматизированная информационно-справочная система (АИСС) «Кадры» | Автоматизация процесса управления кадровым составом правоохранительных органов. | Технологическая привязка к документам личного дела сотрудника. Возможность графического представления исходных данных. Работа в интерактивном режиме. |
| АИСС «Сводка» | Оперативное получение информации о событиях, происшествиях и преступлениях. | Работа с базой данных, создаваемой по поступающей в органы внутренних дел оперативной информации о происшествиях и преступлениях, поиск в БД по реквизитам. |

ФУНКЦИИ ИНТЕГРИРОВАННОЙ ПРОГРАММНОЙ ПЛАТФОРМЫ

Приведенный перечень АИС актуализирует, в частности, задачу поиска баланса между инновациями и практическим удобством использования. Растущее число систем и фрагментированных потоков данных может привести к разрозненности информации, что усложнит совместную работу и принятие решений внутри организации.

Данные Интерпола иллюстрируют масштаб этой проблемы. Интерпол управляет 125 миллионами полицейских досье и ведёт 19 различных баз данных, которые доступны сотрудникам правоохранительных органов стран-участниц. Эти базы данных

ежегодно просматриваются 5,9 миллиарда раз, что приводит к 1,4 миллионам «совпадений» или совмещений, каждое из которых помогает продвинуться в расследовании. Это показывает, какой огромный объём данных приходится обрабатывать правоохранительным органам и насколько велик риск информационной перегрузки [3].

Наиболее оптимально расширить возможности правоохранительной деятельности в современных условиях цифровизации общества, решить указанные проблемы и увеличить эффективность правоохранительной аналитики можно с помощью единой мощной интеллектуальной программной платформы для анализа преступности, которая должна включать функции, приведенные в таблице 2.

Таблица 2

Функции программной платформы

| Функция | Содержание |
|---|---|
| Единый обзор всех оперативных, агентурных, статистических, криминогенных и криминологических данных. | Объединение всех внутренних и внешних данных в единую базу знаний, обеспечивающую комплексное представление информации для принятия точных решений. |
| Расширенные инструменты интеллектуального анализа: - прогнозная аналитика; - геопространственный анализ; - визуализация полученных результатов для улучшения понимания сложных данных. | Комплексная работа с криминологической информацией. Включает анализ финансовых потоков, выявление закономерностей, анализ текста, визуальные запросы, сводки и управление делами. Повышает эффективность выявления горячих точек, закономерностей и тенденций преступности, прогнозирования будущих криминальных инцидентов, стратегического планирования и распределения ресурсов. |
| Поисковые запросы | Возможность использования расширенных поисковых запросов в диалоговом режиме, контекстного поиска с целью выявления преступных схем и быстрого отсеивания подозреваемых из огромных массивов данных. |
| Искусственный интеллект (ИИ) | Интеллектуальные возможности поиска позволяют выявлять скрытые взаимосвязи и связи, необходимые для работы правоохранительных органов. |
| Формирование автоматизированных отчетов | Встроенная автоматизация формирования отчетов различного содержания и сводок, создаваемых с помощью LLM (Large Language Model – большая языковая модель) |
| Возможность обучения | Обучение по категориям пользователей,строенная система повышения квалификации, индивидуализированное и персонифицированное онлайн-обучение. |
| Безопасность | Встроенная система безопасности данных и пользователей. |
| Модульность построения | Обеспечение адаптационной гибкости для различных условий эксплуатации, пользователей и решаемых задач. |

Функции интегрированной программной платформы – объединение данных, прогнозная аналитика, геопространственный анализ, визуализация данных и др. – должны быть тщательно изучены на предмет соответствия программного обеспечения как текущим, так и возможным будущим задачам обеспечения правоохранительной деятельности.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИСПОЛЬЗУЕМЫХ ПЛАТФОРМ

Для принятия решения о соответствии наиболее популярных универсальных платформ выработанным нами требованиям к их функциональным возможностям необходимо провести сравнительный анализ. Ниже приведем описание каждой из платформ.

Платформа Maltego – программное обеспечение, используемое для анализа и криминалистики с открытым исходным кодом, разработанное компанией Paterva из Претории, представляет собой универсальную платформу для проведения расследований. Она обеспечивает:

- мониторинг социальных сетей в режиме реального времени с использованием ИИ-анализа для управления большими объемами данных;
- на основе общедоступных данных выполняет предварительный OSINT-поиск подозреваемых и потенциальных преступников;
- хранение полученных результатов в единой базе в одном формате;
- проведение внутреннего аудита;
- удобный доступ как к внутренним, так и к внешним данным;
- удовлетворение потребностей организации в проведении расследований и соблюдении нормативных требований;
- помочь в расследованиях дел о мошенничестве, киберпреступлениях и многом другом.

Платформа Maltego нашла применение в таких структурах, как Интерпол, TrendMicro, Malformity Labs, 4theONE Foundation, CSITTech, Technoverse Blog, CCP Games. Инструмент подходит как для внутренних расследований, так и для поиска информации в открытых источниках (OSINT, Open source intelligence).

Для доступа ко всем функциям необходима полная лицензионная версия. Пользование платформой представляет сложность для новичков и зависит от доступности трансформов и подключенных источников данных.

Ориентированность на использование в государственных структурах и частных компаниях для

решения задач противодействия киберпреступности, обеспечения кибербезопасности и расследования компьютерных инцидентов ограничивает применение данной платформы для полноценной правоохранительной аналитики.

Платформа Visallo с открытым программным кодом нашла применение в правоохранительных органах, проведении расследований, анализе данных и Big Data. Она позволяет пользователям загружать, интегрировать и анализировать различные данные, чтобы выявлять взаимосвязи и закономерности, при этом поддерживает множество типов данных и форматов визуализации. Ключевые особенности Visallo:

- загрузка и интеграция различных типов и форматов данных, включая текст, изображения, видео, Excel, CSV, JSON и т.д.;
- визуализация данных для обнаружения взаимосвязей с помощью графиков, карт, временных шкал, диаграмм;
- анализ отношений между людьми, местами, событиями, темами и настроениями;
- обеспечение сотрудничества с другими аналитиками с помощью рабочих пространств и встроенного чата;
- расширенная функциональность с помощью открытых API и поддержки онтологий;
- работа с большими массивами данных для выявления закономерностей и аномалий;
- экспорт результатов анализа в виде изображений, PDF-файлов и данных.

Благодаря интуитивно понятному пользовательскому интерфейсу, разработанному специально для проведения расследований, Visallo представляет собой единую визуальную платформу для сбора разрозненных источников данных и анализа сложной информации. Ее используют СМИ, правоохранительные органы, спецслужбы и коммерческие организации для ускорения выявления внутренних угроз, проведения расследований с использованием открытых источников, борьбы с отмыванием денег и в других случаях, когда требуется объединение и визуализация данных.

Платформа Linkurious – это локальное программное решение для визуализации и анализа графов, которое используется для выявления мошенничества, в целях разведки, кибербезопасности или управления ИТ. Она позволяет пользователям находить скрытые связи, выявлять закономерности и сообщества, а также визуализировать сложные взаимосвязи между данными.

Данная платформа представляет собой мощное и простое в использовании программное обеспе-

чение для выявления и расследования, в первую очередь, финансовых преступлений. С её помощью расследовали дело о документах Panama Papers, обнаружив 4,8 млн. файлов, в том числе доказательства отмывания денег, уклонения от налогов или политической коррупции, а также дело о швейцарских утечках, где выявили схему уклонения от налогов, в рамках которой 180,6 млрд. евро прошли через счета Банковской корпорации Гонконга и Шанхая.

Платформа DataWalk предназначена для проведения расследований на основе оперативных данных, результаты которых являются более своевременными, полными и согласованными благодаря сбору всей информации, связанной с любым расследованием, на всех необходимых уровнях для получения полной картины. Кроме того, DataWalk легко подключается к сервисам с открытым исходным кодом и сервисам на основе подиски.

Платформа нашла применение в здравоохранении, телекоммуникациях, розничной торговле и электронной коммерции, государственном и энергетическом секторах, кибербезопасности, консалтинге и бизнес-аналитике, правоохранительных органах. Она позволяет использовать внешние источники информации для расширения существующих данных и, что особенно важно, обращаться к информационным ресурсам МВД России, чтобы получить дополнительную информацию об объекте, а

затем импортировать, отображать и связывать данные для получения более полных представлений и аналитики в рамках текущих исследований или оперативно-розыскных мероприятий.

Программное обеспечение DataWalk преобразует сложные данные в полезную информацию, предоставляя всесторонний обзор всех актуальных данных. С помощью визуальных запросов, контекстного поиска и аналитики на основе искусственного интеллекта возможно значительно упростить анализ оперативно-розыскных, наблюдательных и иных мер, сократив количество задач, выполняемых вручную. Расширенные возможности искусственного интеллекта и графа знаний автоматизируют сбор, подготовку, анализ и составление отчетов, обеспечивая быструю и безопасную совместную работу внутри МВД России и между ведомствами.

Используя искусственный интеллект и граф знаний, программное обеспечение DataWalk Law Enforcement Intelligence объединяет внутреннюю информацию из различных источников, таких как финансовые отчёты, социальные сети и общедоступные записи, автоматизирует импорт, связывание и анализ данных, сокращая объём ручной обработки.

Анализ соответствия указанных платформ сформулированным выше требованиям к функциональным возможностям приведен в таблице 3.

Таблица 3

Универсальные программные платформы и их функции

| Функции | Наименование платформы | | | |
|-----------------------|--|--|---|--|
| | Maltego ¹ | Visallo ² | Linkurious ³ | DataWalk ⁴ |
| Единый обзор данных | ✓ “Всё-в-одном” для расследований; объединение источников через Transform Hub/продукты | ✓ Объединение источников в «Workspace» | ✓ Платформа для унификации и обогащения данных | ✓ Единая корпоративная платформа (One Unified Graph & AI Enterprise Platform), 360-профили |
| Комплексная аналитика | ✓ Графовая/расследовательская аналитика | ✓ Расширенная аналитика расследований | ✓ Графовая аналитика + ИИ | ✓ Графовая и динамическая аналитика, много представлений |
| Прогнозная аналитика | Встроенная предиктивная аналитика не заявлена, но возможна через интеграции/скрипты | ML-плагины/интеграцию. Не ключевая функция | Указаны ИИ/ML, разрешение объекта; предиктивная аналитика, как функция, не заявлена | ✓ ML/LLM/графовые вложения; автоматизация ML-пайплайнов |

¹Maltego. – URL: <https://www.maltego.com/> (Дата обращения: 22.09.2025)

²Visallo. – URL: <https://www.visallo.com/> (Дата обращения: 22.09.2025)

³Linkurious. – URL: <https://linkurious.com/> (Дата обращения: 22.09.2025)

⁴Datawalk . – URL: <https://datawalk.com/> (Дата обращения: 22.09.2025)

Продолжение таблицы 3

| Функции | Наименование платформы | | | |
|---|--|---|--|--|
| | Maltego | Visallo | Linkurious | DataWalk |
| Графическая аналитика | ✓ Да (графы) | ✓ Основная функция платформы | ✓ Да (графы) | ✓ Да (линк-чарты) |
| Геопространственный анализ | ✓ Через гео-трансформы (Google Maps и др.) | ✓ Geo view | ✓ Встроенная «Geography»/geo-mode | ✓ Карты/гео-оверлеи в сценариях |
| Сквозные запросы | Через Transforms/ Machines по источникам; единый движок запросов — ограниченно | ✓ Мульти-хоп запросы в графе | ✓ Конструктор запросов без использования кода (No-code Query Builder), сложные запросы | ✓ No-code мульти-хоп/ контекстный поиск по объединённым наборам |
| Наличие ИИ | ✓ Например, анализ настроек на основе ИИ | Упор на графовую аналитику, ML через расширения | ✓ Графовая аналитика&ИИ, связывание записей, которые относятся к одному реальному объекту в разных источниках данных | ✓ Композитный ИИ, использование больших языковых моделей (LLM) для генерации краткого и информативного резюме, машинное обучение и онто-логика |
| Безопасность | Технология единого входа (Enterprise-SSO)/локальное хранение в дорожной карте | ✓ Контрольные механизмы, аудит | ✓ Управление доступом на основе ролей (Role-Based Access Control, RBAC), модульная аутентификация, совместимость с корпоративной безопасностью | ✓ Гранулярный (к необходимым ресурсам) доступ, аудит, протоколы для аутентификации и управления доступом LDAP/SAML, и др. |
| Комбинированные и автоматизированные отчеты | ✓ PDF-отчёты из графов | Экспорт есть, автоматизация ограничена | Экспорт/публикация; авто-репортинг не заявлен | ✓ Генерация отчётов вручную и автоматически (в т.ч. LLM-саммари) |
| Возможность обучения | ✓ Академия Maltego (обучающие курсы по запросу в реальном времени) | ✓ Обучающие материалы и видео | ✓ Академия Linkurious (обучение и сертификация) | ✓ Обучение и сертификация |
| Модульность построения | ✓ Отдельные продукты (Graph/Monitor/Evidence/ Data) | ✓ Расширяемость через плагины | ✓ Прямо заявлена «модульная платформа» | Единая платформа; модульность не акцентируется (гибкая конфигурация) |

С точки зрения авторов, наиболее оптимальным вариантом использования универсальной программной платформы для реализации информационно-аналитического обеспечения правоохранительной деятельности является платформа DataWalk.

ЗАКЛЮЧЕНИЕ

Цифровая трансформация общества определяет условия осуществления правоохранительной деятельности, которые в актуальных условиях должны быть основаны на интеграции науки о данных и аналитики, как ключевого условия улучшения процессов принятия управленческих решений и предventивных оперативных мероприятий, направленных на борьбу с преступностью.

Интегрирующим элементом внедрения формальной логики, алгоритмов прогнозирования и

предупреждения преступлений, обработки BigData, сквозной и прогнозной аналитики, инструментов решения слабо формализуемых задач, визуализации и распознавания отчетных форм, оптимизации моделей управления в сочетании с машинным обучением являются универсальные программные платформы.

На основании проведенного обзора и анализа можно сделать вывод о том, что лидером по совокупности выполняемых функций является платформа DataWalk благодаря интеграции данных, визуальному поиску, возможностям прогнозной аналитики, более продвинутым инструментам безопасности и другим преимуществам. Таким образом, задача формирования единого информационного пространства правоохранительных органов может максимально эффективно быть решена посредством внедрения и использования указанной комплексной единой программной платформы.

СПИСОК ЛИТЕРАТУРЫ

1. Prayatno C., Tohari M., Susilowati T. The Impact of Using Technology And Innovation In Law Enforcement In The Era Of Digitalization. Jurnal Ekonomi Teknologi dan Bisnis (JETBIS). 2024. Vol. 3. No. 8. P. 1026-1033.
2. Ahmed I. Navigating Ethics And Risk In Artificial Intelligence Applications Within Information Technology: A Systematic Review. American Journal of Advanced Technology and Engineering Solutions. 2025. Vol. 1. No. 01. P. 579-601.
3. Савотченко С.Е., Акапьев В.Л. Информатизация правоохранительной деятельности в условиях цифровой трансформации общества // В сборнике: Развитие информационных технологий органов внутренних дел Российской Федерации.- Москва, 2025. С. 82-91.

УДК: 51, 004.92

Математические аспекты симметрии в изобразительном искусстве

E.S. Goncharenko, I.A. Kirillov

Mathematical Aspects of Symmetry in the Visual Arts

Abstract. This article discusses the mathematical aspects of the concept of symmetry in relation to ornamental art and poetry. The flourishing of ornamental art was achieved in Moorish Spain during the 13th to 15th centuries. The Alhambra Palace served as a showcase of the artistic and mathematical achievements of ancient artists. The article analyzes the influence of Moorish ornaments on the current state of mathematical group theory. Connections are established between the portable symmetry of geometric borders and the rhythm of poetic speech.

Keywords: symmetry, group theory, symmetry groups on a plane, ornament, mosaic, Alhambra.

Явление мавританских орнаментов на современное состояние математической теории групп. Отмечается аналогия симметрии бордюров и поэтической симметрии. Устанавливаются связи переносной симметрии геометрических бордюров и ритма поэтической речи.

Ключевые слова: симметрия, теория групп, группы симметрии на плоскости, орнамент, мозаика, Альгамбра.

СИММЕТРИЯ В ОРНАМЕНТАЛЬНОМ ИСКУССТВЕ

Считается, что геометрический орнамент занимал главное положение в системе искусств ислама и являлся творческим пространством, в котором художники могли реализовывать свои замыслы¹. В исламской культуре по всему мусульманскому миру орнаментальное искусство рассматривалось как кратчайший путь в духовную сферу очищения души и разума.

Коран запрещал изображать живые существа. И весь свой талант, фантазию, и вдохновение, все свои знания и умения арабские художники вкладывали в создание симметричных узоров.

Принципы симметрии в истории человеческой культуры стоят у самых ее истоков. Выдающийся математик XX века Герман Вейль справедливо отмечал: «Симметрия — это идея, с помощью которой человек веками пытался объяснить и создать порядок, красоту и совершенство» [1].

По определению из Википедии: «Симметрия (др.-греч. Συμμετρία = «соподобие»; от συν-

Е.С. Гончаренко¹И.А. Кириллов²

¹Кандидат филологических наук, доцент, Московский государственный лингвистический университет. elisea@rambler.ru

²Кандидат технических наук, доцент, Московский государственный лингвистический университет. I.A.Kirillov@gmail.com

Аннотация. В статье обсуждаются математические аспекты понятия симметрии применительно к орнаментальному искусству и поэзии. Расцвет орнаментального искусства был достигнут в мавританской Испании XIII-XV веков. Дворец Альгамбры явился демонстрацией художественных и математических достижений древних художников. Анализируется влияние мавританских орнаментов на современное состояние математической теории групп. Отмечается аналогия симметрии бордюров и поэтической симметрии. Устанавливаются связи переносной симметрии геометрических бордюров и ритма поэтической речи.

«совместно» + μετρέω «мерю»), в широком смысле — соответствие, неизменность (инвариантность), проявляемые при каких-либо изменениях, преобразованиях (например: положения, энергии, информации, другого)²».

Как показал исторический опыт, исчерпывающее определение симметрии может быть дано только на строгом языке математики и стало возможным в результате создания отдельной математической дисциплины, называемой «Теория групп».

Начнем с некоторых математических определений. Любое взаимно однозначное отображение каждой точки плоскости (точки прообраза) в другую точку той же плоскости (точку образа) называется отображением плоскости на себя. Точка, отображаемая в себя, называется неподвижной. Отображение, при котором все точки плоскости неподвижны, называется тождественным (или нейтральным, или единичным). Для обозначения такого отображения обычно используется буква «E». Отображения, в которых расстояние между любой парой точек-образов остается неизменным и для их точек-образов, называется «отображением движения».

¹ The Evolution of Style | Pattern in Islamic Art. URL: <https://patterninislamicart.com/the-evolution-of-style> (Дата обращения: 31.08.2025).

² Симметрия. URL: <https://ru.wikipedia.org/wiki/Симметрия> (Дата обращения: 31.08.2025).

Наглядным представлением «отображения движения» является как бы создание дубликата исходной плоскости после определенного передвижения, как единого целого, представляющего собой преобразованную плоскость. При этом каждая точка преобразованной плоскости является образом точки исходной плоскости, над которой она оказывается.

СИММЕТРИЯ ЗЕРКАЛЬНОГО ОТРАЖЕНИЯ

Зеркальным отражением плоскости относительно прямой линии называется преобразование, при котором каждая точка M переходит в точку M' такую, что отрезок MM' перпендикулярен этой фиксированной прямой линии на плоскости и делится ею пополам. При этом фиксированная прямая называется осью отражения (осью симметрии). Все точки оси симметрии естественно являются неподвижными точками зеркального отражения. В терминах «отображения движения» плоскости зеркальное отражение сводится к повороту на 180° вокруг оси симметрии.

Для примера, пусть исходная плоскость содержит геометрическую фигуру, представляющую собой равнобедренный (для определенности – не равносторонний) треугольник с выделенной высотой, опущенной на сторону, неравную двум другим. Будем считать, что прямая линия, содержащая выделенную высоту треугольника, является осью симметрии. При зеркальном отражении (при повороте плоскости на 180° вокруг оси симметрии) элементы равнобедренного треугольника полностью совмещаются.

Таким образом, геометрическая фигура равнобедренного треугольника обладает зеркальной симметрией относительно выделенной оси симметрии. Для обозначения рассмотренного преобразования зеркального отражения будем использовать символ M_o (или m_o), где индекс указывает на соответствующую ось симметрии (в случае если ось не единственна). В общем случае, для обозначения взаимно однозначного отображения произвольного множества на себя, сохраняющего неизменным определенные на нем структурные элементы, используется термин «автоморфизм множества» (или, если такое преобразование не единственno, говорят о совокупности «автоморфизмов»).

В рассматриваемом случае у нас существует два автоморфизма плоскости, оставляющих неизменной конфигурацию равнобедренного треугольника. Это тождественное преобразование E , оставляющее неизменным любое преобразуемое множество, и зеркальное преобразование M_o . Именно эти два автоморфизма $\{E, M_o\}$ полностью определяют

симметрию плоскости с равнобедренным треугольником.

Но оказывается, что в определении симметрии важны не только сами автоморфизмы, зависящие от конкретного исходного множества, сохраняющего свою конфигурацию (равнобедренный треугольник), а алгебраическая структура множества автоморфизмов, получившая название «группы симметрии конкретного множества автоморфизмов».

Для задания такой структуры на множестве $G_m = \{E, M_o\}$ вводится внутренняя бинарная ассоциативная операция $*$ (вместо звездочки часто используется знак умножения, да и саму групповую операцию называют «умножением») – правило, по которому каждой паре элементов из G_m ставится в соответствие также некоторый элемент из G_m [2]. Математическая запись этого правила для E и M_o имеет следующий вид: $E * M_o$, а результат операции определяется тем автоморфизмом, который возникнет в результате последовательного применения к исходной плоскости автоморфизмов E и M_o .

В нашем случае E тождественное преобразование плоскости оставляет ее неизменной, а на втором шаге зеркальное отражение M_o повернет исходную плоскость на 180° вокруг оси симметрии (совпадающей с высотой треугольника) так, что результирующим преобразованием останется M_o , что может быть записано в виде: $E * M_o = M_o$.

Аналогичные рассуждения относительно последовательного двукратного зеркального отражения приводят к возвращению исходной плоскости в первоначальное положение. Таким образом, получаем: $M_o * M_o = E$. В довершение имеем: $M_o * E = M_o$ (хотя в общем случае операция $*$ – суперпозиции преобразований и не коммутативна). Наконец, $E * E = E$. Здесь следует заметить, что операцию суперпозиции преобразований в математических текстах принято записывать справа налево, однако этой данью традиции в данном контексте можно пожертвовать без ущерба существу изложения.

В теории групп результаты осуществления бинарных операций принято (бывает удобно) записывать в виде так называемой таблицы Кэли (в первом столбце – первый операнд, в первой строке – второй операнд, на пересечении соответствующей строки и столбца – результат операции суперпозиции). Приведенная Таблица 1 является таблицей Кэли группы зеркального отражения.

Таблица 1

| | | |
|-------|-------|-------|
| * | E | M_o |
| E | E | M_o |
| M_o | M_o | E |

Из этой таблицы легко усматриваются свойства элементов (автоморфизмов) группы симметрии. E – нейтральный элемент, оставляющий любой другой элемент в результате «умножения» без изменения. Для каждого элемента – автоморфизма существует обратный элемент – автоморфизм, «умножение» на который приводит к тождественному преобразованию. Для обозначения элемента, обратного данному, используется обозначение возведения в «минус первую степень». В нашем случае $M_o^{-1} = M_o$, а $E^{-1} = E$. Множество $G_m = \{E, M_o\}$ с внутренней бинарной ассоциативной операцией, с нейтральным относительно этой операции элементом, и наличием у каждого элемента обратного, определяет алгебраическую структуру, называемую группой $G_m = \{E, M_o\}$. Ассоциативность операции в общем случае обеспечивает возможность произвольным образом осуществлять расстановку скобок при наличии более двух операндов.

В довершение разговора о группе автоморфизмов зеркальной симметрии следует упомянуть теорему Кэли [2]. Согласно этой теореме, любая конечная группа изоморфна (простыми словами – «совпадает») вполне определенной «подгруппе симметрической группы». Не вдаваясь в детали математической терминологии, это означает, что группа зеркальной симметрии всегда неизменна, и в этом смысле однозначно описывает саму зеркальную симметрию.

Иными словами, где бы ни возникло зеркальное отражение и соответственно группа зеркальной симметрии, везде её описание однозначно будет определяться соответствующей алгебраической группой, для обозначения которой, как правило, используется символ m_o (индекс при необходимости указывает на заданную ось симметрии, но в очевидных случаях может отсутствовать). Графически ось зеркальной симметрии обозначается сплошной линией, или даже двумя параллельными рядом проведенными линиями.

СИММЕТРИИ ВРАЩЕНИЯ

Перейдем теперь к преобразованиям плоскости, связанным с поворотами вокруг некоторой неподвижной точки (центра вращения). Каждая точка-прообраз, поворачиваясь вокруг неподвижной точки на фиксированный угол вида $\frac{360^\circ}{n}$ (где $n \geq 2$, $n \in \mathbb{Z}$, \mathbb{Z} -множество целых чисел), попадает в точку-образ. Симметрия, связанная с преобразованиями поворота плоскости, носит название поворотной или симметрии вращения.

Для примера пусть исходная плоскость содержит геометрическую фигуру, представляющую собой равносторонний треугольник. На плоскости будем рассматривать поворотные преобразования на углы кратные $\frac{360^\circ}{3} = 120^\circ$, вокруг точки пересечения высот треугольника.

Эти преобразования, оставляя неизменным равносторонний треугольник, являются автоморфизмами, а равносторонний треугольник обладает поворотной симметрией третьего порядка. Точка пересечения высот треугольника является центром симметрии (вращения) 3-го порядка. Для обозначения поворотного на 120° автоморфизма будем использовать символ V_3 (иногда для указания на поворотную симметрию, в частности третьего порядка, бывает достаточно указания числа 3).

Среди других поворотных преобразований рассматриваемой плоскости на углы, кратные 120° , имеются поворотные преобразования на 240° и на 360° . Первое из них может быть представлено в виде двух последовательных поворотов V_3 (суперпозиции): $V_3 * V_3 = V_3^2$. Трехкратный же поворот на 120° приводит к исходному состоянию, являясь, таким образом, тождественным преобразованием, что может быть записано следующим образом: $V_3 * V_3 * V_3 = V_3^3 = E$. Этим самым определена полная циклическая группа собственно поворотной симметрии третьего порядка: $G_v = \{E, V_3, V_3^2\}$ с соответствующей алгебраической структурой, записанной в виде таблицы Кэли (Таблица 2), приведенной ниже.

Таблица 2

| * | E | V_3 | V_3^2 |
|---------|---------|---------|---------|
| E | E | V_3 | V_3^2 |
| V_3 | V_3 | V_3^2 | E |
| V_3^2 | V_3^2 | E | V_3 |

Аналогичным образом может быть построена группа вращения любого натурального порядка.

Группы, состоящие из поворотов на угол, кратный $\alpha = \frac{360^\circ}{n}$, следуя общепринятой терминологии [1, 2], будем называть «циклическими группами собственных вращений». Описанные два типа симметрий: зеркальная и собственно поворотная, могут сочетаться («комбинироваться»), образовывая новый тип симметрии, иногда называемый «поворотно-зеркальной».

Так к собственным вращениям равностороннего треугольника вполне естественно могут быть добавлены зеркальные отражения относительно осей симметрии, проходящих в случае такого тре-

угольника по каждой из трех возможных его высот. Для описания поворотно-зеркальной группы автоморфизмов правильного треугольника приведем ее таблицу Кэли. При этом зеркальные отражения (относительно осей отражения, проходящих через каждую из высот треугольника) будем обозначать соответствующими символами: M_1, M_2, M_3 . Определенная таким образом группа автоморфизмов $G = \{E, V_3, V_3^2, M_1, M_2, M_3\}$ получила название «диэдральная группа третьего порядка» [1]. Таким образом, в диэдральных группах удачно сочетаются симметрии двух типов: собственно вращения и зеркального отражения. Кроме того, оказывается, что «циклические» и «диэдральные» группы являются единственно возможными видами поворотной симметрии на плоскости [2].

Таблица 3

| * | E | V_3 | V_3^2 | M_1 | M_2 | M_3 |
|---------|---------|---------|---------|---------|---------|---------|
| E | E | V_3 | V_3^2 | M_1 | M_2 | M_3 |
| V_3 | V_3 | V_3^2 | E | M_2 | M_3 | M_1 |
| V_3^2 | V_3^2 | E | V_3 | M_3 | M_1 | M_2 |
| M_1 | M_1 | M_3 | M_2 | E | V_3^2 | V_3 |
| M_2 | M_2 | M_1 | M_3 | V_3 | E | V_3^2 |
| M_3 | M_3 | M_2 | M_1 | V_3^2 | V_3 | E |

Из рассмотрения таблицы Кэли диэдральной группы вращений третьего порядка (Таблица 3) читатели, знакомые с элементами теории групп, легко поймут, что верхний левый квадрат 3×3 определяет коммутативную циклическую подгруппу, являющуюся нормальным делителем диэдральной группы. При этом три зеркальных отражения $\{M_1, M_2, M_3\}$ составляют смежный класс диэдральной группы по циклической подгруппе. Кроме того, диэдральная группа содержит еще три подгруппы, соответственно: $G_1 = \{E, M_1\}$, $G_2 = \{E, M_2\}$, $G_3 = \{E, M_3\}$.

На рисунке 1 представлены фигуры, каждая из которых обладает поворотной симметрией 3-го порядка. Причем первая и третья фигуры обладают поворотно-зеркальной симметрией, а вторая фигура – только поворотной симметрией 3-го порядка. Из рассмотрения первой и третьей фигуры еще раз можно убедиться, что как бы ни были различны сами изображения фигур, их группы симметрий совпадают, обеспечивая совпадение их симметрий (поворотно-зеркальных симметрий).

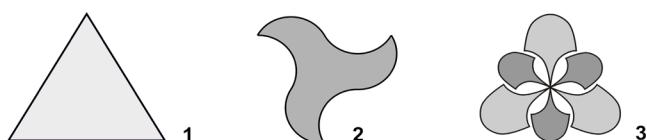


Рис. 1. Примеры фигур, обладающих поворотной симметрией 3-го порядка

Простейшими диэдральными формами симметрии наряду с правильным треугольником обладают правильные многоугольники на плоскости: квадрат, правильный 5-угольник, правильный 6-угольник и т.д.

СИММЕТРИИ ПЕРЕНОСА

Pрассмотрим теперь взаимно однозначные преобразования двумерной плоскости, связанные с параллельными переносами, сдвигами плоскости вдоль линии, совпадающей с направлением фиксированного вектора \bar{T} (оси переноса) на расстояние, кратное длине вектора $t = |\bar{T}|$ (период переноса). Такие преобразования могут привести к так называемым переносным (трансляционным) симметриям. Естественно, такими симметриями могут обладать только бесконечные вдоль оси переноса объекты с периодически повторяемым элементом. Очевидно, изобразить бесконечные объекты не представляется возможным, поэтому при рассмотрении переносных симметрий придется обходиться некоторыми конечными их приближениями, не влияющими на логику изложения. Более того, бесконечные объекты потребуют для своего описания бесконечные группы симметрии, для которых использование таблиц Кэли становится неприемлемым.

В изобразительных искусствах объекты с периодическими узорами вдоль одного выделенного направления получили название «бордюры». Для примера обратимся к рисунку 2, на котором вектор переноса \bar{T} направлен горизонтально вправо, а повторяемый элемент представлен в виде виньетки «■». Автоморфизмы фигуры, изображенной в верхней части рисунка (с поправкой на предполагаемую бесконечность), представляют собой тождественное преобразование E, отображение T (единичный сдвиг плоскости по направлению вектора \bar{T} на расстояние $|\bar{T}|$), обратное отображение T^{-1} (единичный сдвиг против направления вектора \bar{T}) и их последовательные суперпозиции ..., $T^3, T^2, T^1, E, T^1, T^2, T^3, \dots$ Это бесконечное множество автоморфизмов образует бесконечную группу переносной симметрии относительно операции суперпозиции (последовательного выполнения переносов плоскости), изоморфную группе по сложению множества целых чисел Z [4]. Описываемая фигура, инвариантная относительно групповых автоморфизмов, представляющая собой регулярное повторение элементов «■» с неизменным пространственным ритмом, таким образом, является простейшим «орнаментом линейного бордюра» (обладающим симметрией линейного переноса).

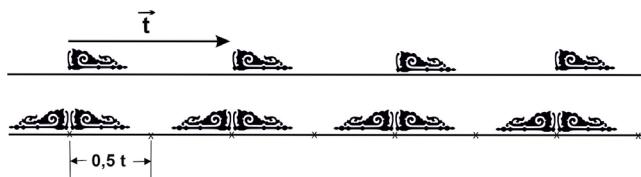


Рис. 2. Примеры переносной симметрии

В простейшем одномерном повторении во времени через равные промежутки-периоды заключается поэтический принцип стихотворного ритма. При этом повторяющиеся промежутки-периоды в стихосложении называются стопами и содержат один ударный слог (обозначаемый \perp) и несколько безударных слогов (обозначаемых \cup). В простейшем случае «ямбическая» стопа содержит один безударный и один ударный слоги $|\cup\perp|$. Ритмическое повторение одинаковых стоп и создает в поэзии ту структуру, которая отличает ее от прозы. Стопы отмеряют условно равные временные интервалы на оси времени поэтической формы, меру стиха. Закон чередования в стихе ударных и безударных слогов называется метром. В стихе метр, число стоп

и их структура определяют стихотворный размер [5]. Например, об «Онегинском» четырехстопном ямбе идет речь, когда в стихотворной строке стоят четыре ямбические стопы. Схематически это может быть представлено «линейным бордюром» следующего вида: $|\cup\perp|\cup\perp|\cup\perp|\cup\perp|$. В данном случае перед нами аналог переносной симметрии вдоль горизонтальной прямой с периодом, равным одной стопе.

Возвращаясь к орнаментальной симметрии линейного переноса необходимо заметить, что она может сочетаться с симметрией отражения.

Это сочетание иллюстрируется изображением нижней части рисунка 2, где центры отражений, расположенные на оси переноса, обозначены значками \times , расстояние между которыми составляет половину периода первоначального переноса $\frac{|t|}{2}$. Кроме того, отражения с переносами могут сочетаться и иным способом, приводя к так называемой «симметрии со скользящей осью». Иллюстрацией сказанного может служить рисунок 3.



Рис. 3. Пример «симметрии со скользящей осью»

На этом рисунке изображена фигура, обладающая переносной симметрией вдоль оси АВ с периодом t . Более того, можно усмотреть, что имеется еще один тип ранее не рассматриваемой симметрии, а именно – симметрия относительно переноса вдоль оси АВ с периодом $\frac{|t|}{2}$ и последующего отражения относительно оси АВ. Ось АВ при этом получила название скользящей оси симметрии с периодом $\frac{|t|}{2}$. Для схематического изображения подобных осей используются не сплошные, а пунктирные прямые линии. Для указания наличия «симметрии со скользящей осью» часто используется символ «g».

Введенная на примере одномерной ситуации симметрия со скользящей осью часто возникает в симметричных построениях и на плоскости.

СЕМНАДЦАТЬ ГРУПП СИММЕТРИЙ ПЛОСКОСТИ

Существенно обогатить сдвиговые симметрии на плоскости позволяет добавление к вектору, определяющему линейный перенос \vec{t} , неколлинеарного ему вектора \vec{s} , и построение на их основе двумерной

периодической структуры плоскости, называемой плоской решеткой. Эта решетка образовывается двумя совокупностями параллельных эквидистантных прямых, направления которых совпадают с направлениями векторов \vec{t} и \vec{s} . Точки пересечения линий называются узлами решетки. Любая пара ближайших параллельных прямых, коллинеарных \vec{t} , при пересечении с любой парой ближайших параллельных прямых, коллинеарных \vec{s} , образуют параллелограмм, именуемый элементарной ячейкой плоской решетки.

Описанная плоская решетка инвариантна относительно любых преобразований сдвигов плоскости вдоль вектора \vec{t} на расстояния, кратные длине этого вектора $|\vec{t}|$, и преобразований сдвигов плоскости вдоль вектора \vec{s} на расстояния, кратные длине этого вектора $|\vec{s}|$. В результате любого переноса каждый узел решетки попадает в другой узел.

В свою очередь, легко видеть, что любым параллелограммом элементарной ячейки может быть задана плоская решетка, а вместе с ней и пространственная симметрия переноса, для обозначения которой часто используется символ «p». Для клас-

сификации симметрий, включающих плоские решетки переноса, принято выделять пять типов таких решеток, получивших название решеток Браве³. В двумерном пространстве каждый из этих пяти типов решеток может быть однозначно описан основными параметрами векторов \vec{t} и \vec{s} :

- a) $\widehat{(\vec{t}, \vec{s})} = 90^\circ$, $|\vec{t}| = |\vec{s}|$ – квадратная решетка;
- b) $\widehat{(\vec{t}, \vec{s})} = 90^\circ$, $|\vec{t}| \neq |\vec{s}|$ – прямоугольная решетка;
- c) $\widehat{(\vec{t}, \vec{s})} = 60^\circ$, $|\vec{t}| = |\vec{s}|$ – гексагональная решетка;
- d) $\widehat{(\vec{t}, \vec{s})} \neq 60^\circ$, $|\vec{t}| = |\vec{s}|$ – ромбическая решетка;
- e) $\widehat{(\vec{t}, \vec{s})} \neq 60^\circ$, $|\vec{t}| \neq |\vec{s}|$ – косая решетка.

В простейшем случае орнамент на плоскости характеризуется только переносной симметрией. Таков, например, орнамент на рисунке 3, слева (построенный на основе косой решетки). Чтобы построить этот орнамент, элементарная ячейка (соответствующий параллелограмм) выбранной решетки была заполнена рисунком, лишенным какой-либо симметрии. Затем эта ячейка многократно повторялась за счет переносов без изменения ее ориентации. Сама элементарная ячейка схематически представлена отдельно на том же рисунке 4, рядом с полученным орнаментом.

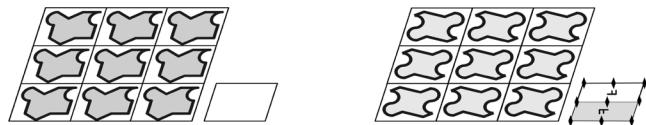


Рис. 4. Два варианта плоского орнамента на основе «косой» решетки

Ранее было отмечено, что симметрия линейного переноса вдоль прямой может сочетаться с симметриями отражений. В случае переносов на плоскости сочетание этих симметрий продолжает иметь место. Что же касается сочетания симметрий переносов с симметриями вращений, то здесь возникает ряд ограничений, связанных с тем, что преобразование вращения решетки естественно должно переводить плоскостную решетку в себя. Тем не менее может быть доказано [4], что значения порядков вращения для сочетаемости со сдвигами должно быть ограничено числами: 2, 3, 4, 6.

Для схематичного обозначения соответствующих осей вращения используются следующие значки: \blacktriangleleft – для оси второго порядка; \blacktriangleright – для оси третьего порядка; \blacksquare – для оси четвертого порядка; \blacklozenge – для оси шестого порядка. Кроме того, каждая из пяти решеток Браве связана с конкретными ограничениями на сочетаемость с другими описанными выше элементарными симметриями (отражения, вращения и отражения со скользящей осью). Самой «бедной» (или самой «требовательной») в смысле сочетаемости является «косая» решетка, единственно допускающая сочетаемость только с симметрией вращения второго порядка.

В подтверждение сказанному, на рисунке 4, слева, изображен пример косой решетки, соответствующий лишь трансляционной симметрии. На этом же рисунке 4, справа, представлена трансляционная симметрия в сочетании с поворотной симметрией второго порядка. Соответственно схематически для левого изображения приводится ячейка, не обладающая какой-либо дополнительной симметрией (эта ячейка пуста). Схема же ячейки косой решетки с комбинацией вращения второго порядка содержит обозначения всех возможных поворотных осей симметрии второго порядка. Кроме того, светлая верхняя половина этой ячейки показывает ту ее часть, которую достаточно заполнить исходным рисунком для создания орнамента на всей двумерной плоскости. Так как при наличии поворотной симметрии второго порядка этот исходный рисунок, часто называемый «основной мотив», повернувшись на 180° , отобразится с верхней и на нижнюю часть ячейки, а вся ячейка затем будет «растранслирована» по всей плоскости.

В общем случае «основной мотив» приводится во взаимодействие с элементами симметрии (осиами поворотной симметрии, осями зеркального отражения, осями трансляции и проч.). Для схематического обозначения «основного мотива» элементарной ячейки часто используется повернутый на 90° символ F .

Кроме схематического изображения элементарной ячейки соответствующей решетки, для обозначения «комбинированных» симметрий плоскости часто используется так называемая нотационная система Германа – Могена [4], упрощенный вариант которой состоит из символов, соответствующих сочетающихся элементарных составляющих симметрий: **p**, **m**, **g** и **число**, обозначающее максимальный порядок симметрии вращения.

Таким образом, рассмотренные выше на рисунке 4 орнаменты могут быть обозначены соответственно **p1** (вращения отсутствуют) и **p2** (вращения на 180°).

С учетом сочетаемости элементарных симметрий оказалось, что количество различных «комбинированных» групп симметрий на плоскости рав-

³ Решетка Браве. URL: https://en.wikipedia.org/wiki/Bravais_lattice. (Дата обращения: 31.08.2025).

но 17. Доказательство этого нетривиального факта было впервые проведено российским кристаллографом и минералогом Евграфом Степановичем Фёдоровым в 1891 году.

ОРНАМЕНТЫ АЛЬГАМБРЫ

Как подчеркивал Герман Вейль [1], образцы всех 17 групп симметрии на плоскости были обнаружены среди декоративных узоров арабских орнаментов. Прежде всего это относится к архитектурным и художественным шедеврам дворцов Альгамбры, создатели которых в совершенстве владели всеми типами симметрии на плоскости.

Альгамбра — архитектурный ансамбль в восточной части города Гранада (Южная Испания). Основ-

ное развитие Альгамбры получила после упадка Халифата Кордовы (в мавританской Испании) с 1230 до 1492, когда Гранада являлась столицей Гранадского эмирата на Иберийском полуострове, а Альгамбра — резиденцией гранадских эмиров⁴.

На облицовке стен многих дворцовых помещений Альгамбры обыгрываются цветные варианты орнамента, приведенного на рисунке 5. На том же рисунке приведено схематическое изображение основного мотива элементарной ячейки соответствующей гексагональной решетки Браве. Для приведенного цветового решения (градации серого) нотационная система Германа — Могена обозначает группу симметрий — **p3** (развернутая запись, указывающая на три типа осей симметрий вращения третьего порядка, выглядит как — **p3(333)**).

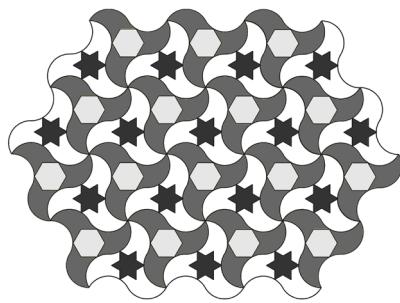
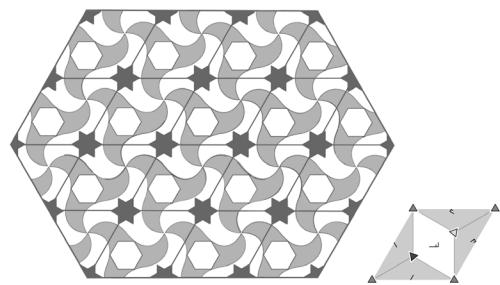


Рис. 5. Серые (8 бит) варианты фрагментов орнамента Альгамбры на основе гексагональной решетки Браве

На рисунке 6 в авторской реконструкции представлен ряд орнаментов дворцов Альгамбры, сохраняющих свою прелест даже в серых цветах

(градации серого использованы в журнальном варианте соответствующих изображений).

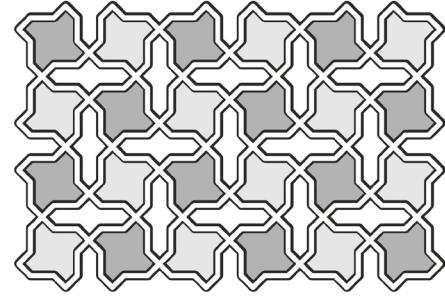
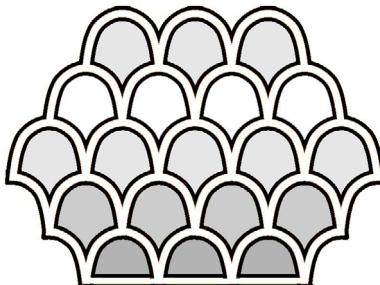
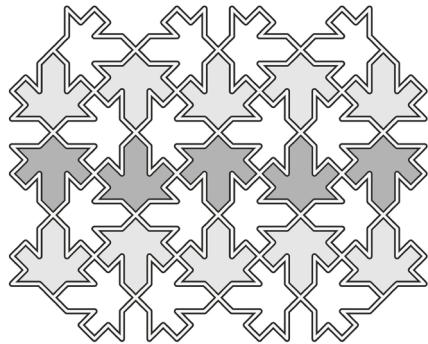


Рис. 6. Варианты фрагментов цветных орнаментов Альгамбры (приведенных в серых восьмибитных тонах) на основе квадратной и ромбической решеток

Обозначение первого из трех орнаментов имеет вид — **p4g** (квадратная решетка), для второго орнамента — **c_m** (символ с используется только для ромбических решеток при наличии зеркальных и скользящих симметрий) и для третьего орнамента — **p4m** (квадратная решетка).

Наряду с практическим освоением всех симметрий на плоскости арабские мастера Альгамбры очень часто в своих потрясающих творениях сочетали их с техниками мозаик. Мозаика — регулярное разбиение плоскости (отвечающее правилам переносной симметрии) набором симметричных

⁴ Альгамбра. URL:<https://ru.wikipedia.org/wiki/Альгамбра> (Дата обращения: 31.08.2025).

замкнутых фигур, которыми можно замостить плоскость без пересечений фигур и без щелей между ними. В простейших мозаиках используются квадраты или прямоугольники. Но и здесь фантазиям мастеров Альгамбры не было пределов.

На рисунке 7 приведены еще три орнаментальных шедевра Альгамбры. Первый орнамент с гексагональной решеткой Браве с симметрией вращения шестого порядка и зеркальными отражениями имеет обозначение **p6m**. Во втором и третьем орнаментах имеет место симметрия вращения четвертого порядка. Более того, можно утверждать наличие

диэдральной группы вращения. Второй орнамент, естественно, аппроксимируется симметрией сдвига с квадратной решеткой Браве. Две ортогональные оси зеркального отражения в нотации Германа – Могена приводят к обозначению **pmm**. С определенной оговоркой третий орнамент также может считаться обладающим переносной симметрией, при этом рассматриваемый орнамент имеет квадратную решетку и, с учетом диагональных осей зеркального отражения и осей переносной симметрии, его группа симметрии будет обозначаться как **p4g**.

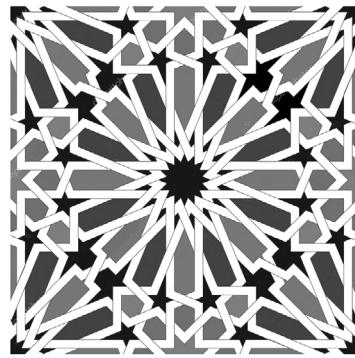
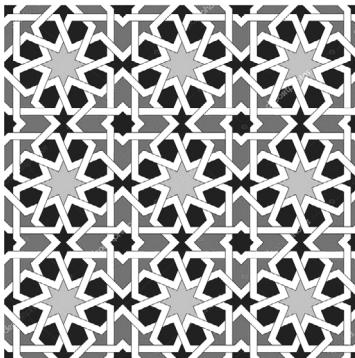
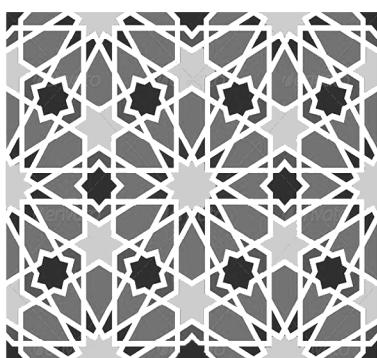


Рис. 7. Три орнаментальных шедевра Альгамбры

ПОСЛЕСЛОВИЕ

Многообразие орнаментальных аналоговых изображений укладывается в строгие математические рамки современной теории групп и может быть описано одной из семнадцати групп симметрии изображений на плоскости. Каждый конкретный тип плоскостной симметрии «компонуется» из основных «элементарных симметрий»: зеркальной, поворотной и переносной. Количество разновидностей элементарных симметрий, пригодных для «компоновки», ограничено четырьмя возможными вращениями и пятью допустимыми типами решеток Браве. Тем не менее, установить нетривиальный факт того, что общее количество всевозможных «комбинированных» групп симметрий на плоскости равно 17, первым в мире удалось нашему соотечественнику кристаллографу, минералогу и математику Е.С. Фёдорову только в конце 19 века. Впервые строгое математическое доказательство этого факта на основе теории групп было получено в 1924 году венгерским математиком Дьёрдем Пойа⁵.

Задолго до упомянутых выше доказательств мавританские мастера не только были знакомы с каждым из 17 возможных типов симметрии на плоскости, но и практически использовали каждый из них в создании своих орнаментальных шедевров, по крайней мере, с начала возведения дворцового комплекса Альгамбры в 1230 году. С этой датой связано начало правления последней мавританской династии Насридов. Как считают многие историки и искусствоведы, именно этот период характеризуется культурным расцветом эмирата Гранады⁶, который являлся прибежищем всех мусульман с отвоеванных у них территорий пиренейского полуострова. В Гранадском эмирата продолжали развиваться великие традиции мавританской поэзии, декоративного искусства и архитектуры. Это продолжалось вплоть до полного завершения реконкисты католическими королевствами Кастилии и Арагона в 1492 году.

Альгамбра и мавританское искусство и поныне продолжают восхищать своими орнаментальными шедеврами, основанными на семнадцати возможных группах плоских симметрий.

⁵ A wallpaper group. URL: https://en.wikipedia.org/wiki/Wallpaper_group?ysclid=mefwnimcwk693236272 (Дата обращения 31.08.2025).

⁶ Гранадский эмир – Википедия. URL: https://ru.wikipedia.org/wiki/Гранадский_эмир (Дата обращения: 31.08.2025).

СПИСОК ЛИТЕРАТУРЫ

1. Вейль Г. Симметрия. Перевод с английского Б.В.Бирюкова и Ю.А.Данилова под редакцией Б.А.Розенфельда. – Москва: Наука, 1968. 192 с.
2. Теория групп. А. Г. Курош. – Москва: Физматлит, 2011. 805 с.
3. Поклонский Н. А. Точечные группы симметрии: Учебное пособие. – Минск: БГУ, 2003. 222 с.
4. Тарасов Л. В. Этот удивительно симметричный мир: Пособие для учащихся. – М.: Просвещение, 1982. 176 с.
5. Гончаренко Е. С., Кириллов И. А. Принципы акцентуации как фундаментальная проблема для понимания ритма испанского стихотворного текста и его информативности // Сборник трудов II Международной научно-практической конференции имени С. Ф. Гончаренко «Высокое искусство перевода», 2022. С.74-85.

От редакции

Михаил Евгеньевич Масленников – один из выдающихся российских криптографов с мировым именем и авторитетом. Читателям хорошо известна его книга «Криптография и свобода», доступная в настоящее время на сервисе Литрес [1].

Наша редакция предоставляла ему слово на страницах журнала, поскольку его мнение всегда остро, принципиально и актуально. Не во всем можно согласиться с Михаилом в плане критики официальной криптографии, но в его мнении всегда есть рациональное зерно и мысли, к которым необходимо прислушаться.

Часто высокие начальники в условиях собственных неприятных провалов, похожих на размещение российских финансовых активов за рубежом, говорят экспертам: «А что же вы не предупреждали, товарищи учёные!?». А когда учёные и эксперты говорят, что они как раз предупреждали, отвечают: «Значит, вы меня не убедили и не настояли на своем мнении!».

Чтобы поводов говорить такое у чиновников от науки было меньше, мы с удовольствием публикуем статью Михаила Евгеньевича на актуальную тему.

Россия и RSA. Почувствуйте разницу

Предисловие

С послевоенных времен до середины 70-х гг. ХХ в. СССР всегда успешно конкурировал с США в области криптографии. Но американская революция в криптографии, связанная с появлением асимметричных криптографических систем, вызвала у советских криптографов некоторое беспокойство. Как ответить американцам на «научном поле»? Думали долго и примерно через 10 лет пришли к выводам, что просто отрицать эту революцию нельзя.

Американцы не спешили. Сам научно-технический прогресс во всем мире, появление компьютерных сетей и Интернета, настоятельно требовали "гражданской" криптографии. И в середине 90-х американцы, наконец, решились: делаем криптографию общедоступной! Этот вывод появился в результате тщательного взвешивания всех «за» и «против», которые провел весьма представительный американский Комитет по изучению национальной политики в области криптографии¹.

Таким образом, между появлением асимметричных криптографических систем с открытыми ключами и до американского вывода об общедоступности криптографии прошло 20 лет. Если бы СССР, а затем Россия смогли "вклиниваться" в этот 20-летний зазор и попытаться завоевать какую-то часть мирового рынка криптографической продукции, то современный мир выглядел бы иначе. Самый удобный момент для этого был именно в начале 90-х.

На мой взгляд, в конце 1992 года был переломный момент, когда российская криптография могла повернуться в ту или иную сторону. С одной стороны, был шанс развития гражданской и свободной криптографии раньше американцев, а с другой – возможность еще круче завернуть гайки. Выбрали второй путь. Ради чего и что это дало? Постараюсь высказать на эту тему свою точку зрения.

Американская революция в криптографии

Давайте подробнее поговорим об американской революции в криптографии. Еще Сталин отмечал американскую деловитость. И эта деловитость неизбежно вела к спорам: а почему криптография должна быть предназначена только для военных и правительственные нужд? Можно ли использовать криптографию, например, в коммерции, именуя такую криптографию гражданской? Тут надо заметить, что никакой коммерции в 70-х годах в СССР официально не было, следовательно, в СССР не было даже намеков на «гражданскую криптографию».

В 70-х годах была доктрина: все, что связано с криптографией, секретно. В 1975 году при посвящении в специальность слушателям IV факультета Высшей школы КГБ говорили: «Никто, ни мать, ни отец, ни жена – никто не должен знать, что вы занимаетесь криптографией».

Американцы же в 70-х годах в первую очередь связывали криптографию с деньгами. И не только.

¹ Комитет по изучению национальной политики в области криптографии, сформирован в США по просьбе Конгресса в ноябре 1993 года Советом по компьютерным наукам и телекоммуникациям Национального исследовательского совета (CSTB).

Бурная научно-техническая революция на западе тоже требовала гражданской криптографии, требовала отбросить устаревшие криптографические догмы о тотальной секретности. Этому способствовало также и стремительное развитие самой криптографии, методов построения надежных криптографических схем и методов их анализа. Со временем взлома немецкой «Энигмы» прошел уже не один десяток лет, криптографы научились создавать машинные (механизированные) шифры высокой надежности.

Особенно сильно со временем «Энигмы» изменилась элементная база, на основе которой создавались машинные шифры. Уже не за горами была тотальная компьютеризация, дававшая возможности реализовывать практически любые криптографические фантазии. Хоть гарантированная стойкость и не являлась чисто математическим понятием, как абсолютная стойкость по Шенону, но постепенно переходила на общепринятый язык: «сделать ничего нельзя, сколько ни пытались, только тотальный перебор ключей».

В американской революции в криптографии все было сделано не по-советски: не было штаба революции, не было революционно настроенных матросов и криков «ура», не было партии и вождя, не было и много другого, что во времена СССР непременно связывали с понятием «революция». Я бы употребил здесь свое любимое слово: нетривиальная революция. Хорошие специалисты, нетривиально мыслящие люди тихо и спокойно совершили переворот в криптографии, которым в полной мере воспользовалось следующее молодое поколение эпохи Интернета.

Критически настроенный читатель непременно заметит: разрешили американцы использовать криптографию не только для военных и правительственные линий связи, но и в коммерции. Так что же здесь революционного и нетривиального? Вполне здравая и разумная идея, все верно. Но с одним непременным условием: к этой идеи надо пристроить «коммерческие ноги», т.е. что-то такое, что позволит криптографии успешно работать в коммерции. Такими «ногами» стала математически великолепная и нетривиальная идея «открытых ключей».

Что самое главное и самое сложное при организации криптографической защиты? Конечно же, ключевая система. Ключи, с помощью которых осуществляется шифрование, должны доставляться адресатам и храниться ими с использованием повышенных мер секретности, ибо компрометация

ключа шифрованной связи равносильна потере секретности. Гарантировать безопасность секретной связи при компрометации ключа к машинному шифру уже нельзя. Криптографы при анализе спрашивают предполагают, что потенциальный злоумышленник знает о шифре все, кроме ключа².

Для хорошей ключевой системы нужна строгая система обеспечения ее безопасности, нужны гарантии отсутствия компрометации ключа перед тем, как он будет использоваться в связи. До середины 70-х годов такие гарантии обеспечивались исключительно организационно-административными мерами, что, естественно, делало секретную связь дорогой и, если так можно сказать, экзотической. Не для простых людей и даже не для коммерсантов, которые, несмотря на все преимущества закрытой связи, реально широко использовать ее не могли.

Первоначально само словосочетание «открытый ключ» воспринималось как какая-то несуразность, вроде «сухой воды». Обычным людям, в том числе и хорошим военным, всегда казалось, что ключ потому и называется ключом, что он должен храниться в секрете, иначе какой-же это ключ?

Что необходимо для появления какой-то нетривиальной идеи? Хорошие специалисты, любящие раскованность и нетривиальность мышления. Но этого мало. Опять вспоминая ленинское «жить в обществе и быть свободным от общества нельзя», для появления нетривиальной, нестандартной и на первый взгляд парадоксальной идеи в обществе должна созреть потребность в ней. В США в середине 70-х годов созрела и все сильнее ощущалась потребность в гражданской криптографии, а следовательно – в такой ключевой системе, которая была бы удобна при использовании в коммерческой сфере. В СССР в середине 70-х годов какая-то минимальная потребность в гражданской криптографии, возможно, существовала, но она была до основания «замурована» различными идеологическими догмами. Тут Ленин оказался прав.

В этой статье я не буду вдаваться в технические подробности всех нетривиальных американских идей, связанных с гражданской криптографией. Полагаю, что интересующийся читатель без труда сможет отыскать их в Интернете, например в Википедии, по некоторым ключевым словам. Моя задача – указать эти ключевые слова. На мой взгляд, наиболее популярным является словосочетание «открытый ключ», который также иногда называют «публичным ключом». С ним тесно связаны

² Правило Кирхгофа (принцип Кирхгофа) в криптографии гласит, что безопасность криптосистемы должна основываться на секретности ключа, а не на секретности алгоритма. Этот принцип был сформулирован в 1883 году голландским криптографом Августом Кирхгофом и до сих пор является основополагающим в криптографии.

«криптографическая система с открытым распределением ключей», «асимметрическая криптографическая схема», где ключом является пара, состоящая из секретного ключа и соответствующего ему открытого ключа. Альтернативой асимметрической схеме является традиционная «симметрическая схема» только с секретными ключами. Симметрическая схема еще пригодится, забывать ее не стоит.

Как появляются нетривиальные идеи

Как же появлялись на свет эти нетривиальные криптографические идеи? Неожиданно, непредсказуемо, ведь общество, в котором хороший специалист пытается получить нетривиальный результат, может либо способствовать, либо мешать этому, но конечный результат зависит от специалиста, от его дара предвидения и веры в успех.

Современный Интернет немыслим без асимметричного алгоритма RSA. RSA – первые буквы фамилий авторов: Rivest, Shamir, Adleman. Вот что написано в Википедии о том, как появилась на свет идея этого алгоритма.

«Рон Ривест, Ади Шамир и Леонард Адлеман из Массачусетского технологического института в течение года предприняли несколько попыток создать одностороннюю функцию, которую было бы трудно инвертировать. Ривест и Шамир, будучи компьютерными учеными, предложили множество потенциальных функций, а Адлеман, будучи математиком, отвечал за поиск их слабых мест. Они опробовали множество подходов, включая "ранцевый" и "перестановочные полиномы". Какое-то время они думали, что то, чего они хотели достичь, невозможно из-за противоречивых требований. В апреле 1977 года они провели Песах в доме одного из студентов и выпили много манишевицкого вина, а затем вернулись к себе домой около полуночи. Ривест, не в силах заснуть, лег на диван с учебником математики и начал думать о своей односторонней функции. Остаток ночи он провел, формализуя свою идею, и к рассвету большая часть статьи была готова. Алгоритм теперь известен как RSA – инициаторы их фамилий в том же порядке, что и в их статье.»

В ноябре того же 1977 года советские криптографы отмечали 60-летие Великой Октябрьской социалистической революции. Подробности опускаем, они интуитивно ясны и ничем не хуже аналогичных подробностей накануне открытия RSA. А вот дальше – расхождение. Ривест остаток ночи провел, формализуя свою идею, а советским криптографам надо было вставать в 5 утра и идти в оперативный наряд перекрывать подходы к Красной площади перед демонстрацией трудящихся.

Советские криптографы не получили в этом оперативном наряде нетривиальных результатов, таких как алгоритм RSA, но в конце наряда наш старшина курса молодцевато и браво козырнул медленно проезжавшей мимо цепочки криптографов правительенной машине, в которой сидел М.А. Суслов. Весь наряд получил за это благодарность.

Начало революции

«Есть у революции начало, нет у революции конца» – говорилось в одной популярной в 70-е годы советской песне. На мой взгляд, для американской революции в криптографии это верно. Ну, а как же отреагировала на ее начало советская криптографическая наука?

В 1976 году Уитфилд Диффи и Мартин Хеллман опубликовали работу «Новые направления в современной криптографии», в которой предложили метод получения секретных ключей, используя открытый канал связи. Этот метод использовал одностороннюю функцию возведения в степень в конечном поле.

Любящие конкуренцию американцы на методе Диффи-Хеллмана (DH) не успокоились и в 1977 году придумали еще и упоминавшийся выше метод RSA. Чем они отличаются, что лучше, а что хуже – мы об этом еще поговорим ниже.

Итак, будем считать началом американской революции в криптографии 1976 год и легко объяснимую красивую систему с открытым распределением ключей, использующую одностороннюю функцию возведения в степень в конечном поле. Но одной лишь асимметричной криптографии для полного удовлетворения потребностей гражданского общества недостаточно. Дело в том, что скорость шифрования асимметрической схемы относительно низкая, намного ниже, чем у симметрической. Но решить эту задачу уже намного проще. С помощью асимметрической схемы шифруется только секретный ключ для симметрической схемы, а дальше основная информация быстро шифруется симметрической схемой.

Таким образом, гражданскому обществу нужна еще общедоступная симметрическая схема, стойкость которой не вызывает сомнений. Про эту схему просто по определению можно сказать, что потенциальный злоумышленник знает все, кроме ключа. В этих условиях она должна обеспечивать гарантированную стойкость.

В 1977 году правительством США был утвержден DES – Data Encryption Standard. Симметрическая схема, с которой сняты все грифы секретности. Попутно заметим, что разработкой DES занималась компа-

ния IBM. Сенат США в 1978 году проверил действия АНБ и признал, что «представители АНБ никогда не вмешивались в разработку алгоритма DES».

Эти три события – появление DH, RSA и DES – произошли почти синхронно, удачно дополняя друг друга. Математикам во всем мире становилось понятно, что в США за создание и использование гражданской криптографии взялись всерьез.

А оно нам надо?

«Оно нам надо?» – так сначала отреагировало советское криптографическое начальство на DH и RSA. Американцы придумали что-то непонятное, как оценить стойкость – неясно, опыта исследований в области асимметричной криптографии нет, потребности в гражданской криптографии – тоже. Принимаем к сведению, однако тратить большие ресурсы на какую-то экзотику не будем. Но математики-криптографы, хорошие специалисты, не могли не оценить красоту и практическую значимость DH и RSA и постепенно стали в них разбираться.

А вот что касается DES, вопрос «Оно нам надо?» вполне уместен. DES – обычная симметричная схема, а в СССР вся спецсвязь на симметричных схемах работает. Почему американцы взяли и в открытую опубликовали свою симметричную схему?

Хорошо запомнившаяся мне сцена конца 1979 года. В отдел, в который я попал после окончания IV факультета, приехал заместитель начальника Главного управления Владимир Николаевич Сачков. Собрали совещание, на котором Сачков обратился ко всему отделу:

– Внимательно проанализируйте американский стандарт DES. Если удастся его взломать, то я буду докладывать об этом очень высоко.

Но высокий доклад так и не состоялся.

Над советскими криптографическими начальниками стояли более могущественные начальники, которые пользовались закрытой связью. Их логика была простой и понятной. Американцы в открытую опубликовали свою симметричную схему шифрования потому, что уверены в ее стойкости. Ведь открытая публикация означает, что анализировать эту схему будут все, кому не лень, и по всему миру. Любой прокол или промах обязательно будет обнаружены, со всех сторон изучены, и информация об этом обязательно будет распространена тоже по всему миру. По меньшей мере, угроза потери репутации. Следовательно, американцы настолько уверены в гарантированной стойкости DES, что не боятся его открытой публикации.

А в СССР? Для чего нужна такая тотальная секретность в криптографии и что за ней скрывается?

Есть ли в СССР такая криптосхема, которую можно было бы отдать на растерзание всему миру? А что, если её нет? Если вся советская секретность служит лишь для того, чтобы скрыть истинное состояние криптографии в СССР? А криптография – это закрытая правительенная связь, это система управления баллистическими ракетами и многое, многое другое, где верить на слово нельзя. Нужны простые и понятные доказательства того, что в СССР криптография не менее сильная, чем в США. Американцы сделали открытый стандарт DES – пусть теперь и советские криптографы ответят на него тем же.

На мой взгляд, подобная логика преобладала в конце 70-х годов. Но об одном можно сказать уверенно: ни о какой гражданской криптографии в СССР в это время речи не шло.

Советский стандарт создавали около 10 лет, он появился на свет только в 1989 году. В результате стандарт получился похожим на DES. Для чего делали советский ГОСТ 28147-89, кому он был нужен в 1989 году? Сколько стоило такой стандарт разработать и – главное – утвердить?

За те 10 лет создания советского стандарта симметричного шифрования произошло много событий. Это и «пятилетка пышных похорон», и сильное падение цен на нефть, это перестройка, гласность, ускорение и многое другое.

Долгая и малопонятная для большинства людей война в Афганистане также длилась почти те же 10 лет, а на войне роль шифрованной связи сильно возрастает, часто от нее напрямую зависят жизни солдат. В низовом армейском звене с давних времен было принято использовать переговорные таблицы. Если, к примеру, командир хочет послать своим подчиненным какой-то приказ, то он должен был взять огромную книгу с переговорными таблицами, закодировать по этой книге свой приказ и послать его, часто по обычному радиоканалу, в закодированном виде. Долго, трудно и неудобно. В реальной боевой обстановке такие переговорные таблицы игнорировались и приказ посыпался в открытом виде.

Что можно было придумать вместо этих доисторических переговорных таблиц? Защищенную портативную радиостанцию. На западе с его научно-техническим прогрессом такие радиостанции уже начали появляться. Вот где пригодились открытые ключи и асимметричная криптография. В СССР стали пристальнее присматриваться к DH и RSA. Они нужны, оказывается, не только коммерсантам, но и военным! Так в чем же проблема?

В элементной базе, точнее – в ее отсутствии. Для реализации новых американских идей в криптогра-

фии нужна была и новая элементная база, микропроцессоры, миниатюрные интегральные микросхемы, различные преобразователи, специальные языки программирования и многое другое. Советская элементная база по старинке состояла из полупроводников, транзисторов, диодов, сопротивлений, конденсаторов, проводов и индукционных катушек. Компьютеры, собранные на советской элементной базе, занимали целые комнаты, были громоздкими, ненадежными, часто ломались. В общем – полностью соответствовали советскому тезису, что «кибернетика – это буржуазная лженаука».

Эпоха 80-х

Пора поговорить о 80-х годах прошлого века и об их влиянии на советскую криптографию. На мой взгляд, в это время советская криптографическая синусоида сделала последнее движение вверх, чтобы затем надолго упасть вниз.

Первая реакция на американские открытые ключи прошла. Стало ясно, что эта американская нетривиальная революция в криптографии сулит значительные выгоды, но и требует современной элементной базы, которой в СССР в то время не было. Появились самые фантастические идеи. Например, создать свой «свечной заводик» для производства специальных микросхем под различные криптографические задачи, в том числе и под асимметричную криптографию. Это было где-то в первой половине 80-х.

Но все-таки основным событием 80-х годов в советской криптографии было появление персональных компьютеров. Вот она, та самая современная элементная база, которой так не хватало не только для реализации новых американских идей, но и при повседневной и будничной криптографии, связанной с анализом традиционных симметричных криптосхем для советской спецсвязи. Первые ПК произвели фурор, работа криптографа-аналитика сразу приобрела иной характер. Быстро стала понятной абсурдность собственного «свечного заводика», который просто по определению не мог составить ни малейшей конкуренции даже самым первым персональным компьютерам IBM PC XT. Появление в СССР персональных компьютеров совершенно неожиданно привело к целому ряду стихийных процессов, к которым советское криптографическое начальство оказалось не готовым.

Но вот руководителем СССР стал Михаил Сергеевич Горбачев и постепенно различные советские НИИ и КБ стали напрягаться. Какую продукцию мы производим? Конкурентоспособна ли она? А что, если завтра всех разгонят? Не пора ли готовить «з-

пасной аэродром» и осваивать какую-то иную, более востребованную профессию?

Кому нужен математик-криптограф, всю свою сознательную жизнь занимавшийся закрытыми и непонятными работами? Вопрос риторический. А вот математик, освоивший персональный компьютер и научившийся на нем программировать – это уже совсем другое дело. Тем более, что сделать из математика программиста сравнительно легко, а получить из программиста хорошего математика часто бывает намного сложнее. Вот такая одноправленная функция.

И как-то чаще стала появляться идея гражданской криптографии, т.е. более широкого использования накопленного криптографического опыта простыми людьми в повседневной жизни. На американскую нетривиальную революцию советские криптографы стали смотреть уже по-другому: вот он, «запасной аэродром»!

Конец 80-х. В СССР персональные компьютеры еще в диковинку, но и в западном мире, несмотря на бурный научно-технический прогресс, доминирующая в то время операционная система MS DOS еще не добралась до криптографии. Потребность в криптографии есть, но хороших криптографических программ под MS DOS мало. Как показала дальнейшая история развития компьютерной техники, такая ситуация сохранится еще примерно 10 лет до конца 90-х. В 1993 г. американцы создадут специальный комитет из представителей различных сфер: науки, производства, бизнеса, государственных и военных ведомств, для выработки новой государственной политики в области криптографии. Он назывался «Комитет по изучению национальной политики в области криптографии».

В 1996 г. в США был опубликован совместный отчет Комитета по изучению национальной политики в области криптографии, Совета по компьютерным наукам и телекоммуникациям, Комиссии по физическим наукам, математике и приложениям и Национального исследовательского совета. Отчет называется «Роль криптографии в обеспечении безопасности информационного общества», он доступен в Интернете по адресу <https://nap.nationalacademies.org/catalog/5131/cryptographys-role-in-securig-the-information-society>.

Комитет пришел к выводу, что преимущества широкого распространения криптографии перевешивают его недостатки, и призвал к изменению национальной политики в криптографии.

Цели новой национальной политики в области криптографии:

- широкая доступность криптографии для всех

законных элементов американского общества;

– дальнейший экономический рост и лидерство ключевых отраслей промышленности США, в том числе (но не только) производства компьютеров, программного обеспечения и средств связи, в глобальной экономике;

– общественная безопасность и защита от внешних и внутренних угроз.

Таким образом в США официально признали, что «преимущества широкого распространения криптографии перевешивают его недостатки», но это произошло только в 1996 году. В конце 1990-х – начале 2000-х гг. появляется открытый пакет криптографических процедур OpenSSL, а в самой популярной мировой операционной системе – Windows – появился так называемый Cryptography Service Provider, динамическая библиотека, содержащая реализации основных криптографических процедур, необходимых для реализации электронной подписи и шифрования.

К американским научно-техническим достижениям в области гражданской криптографии, достигнутым в середине 70-х гг., в середине 90-х гг. добавилась поддержка на государственном уровне «широкого распространения криптографии». Терпеть дальше таинственность и оторванность от реальной жизни профессиональной и квалифицированной криптографии было нельзя в первую очередь из-за стремительного развития компьютерных сетей и Интернета. Между первым появлением блестящих американских идей, связанных с асимметричной криптографией, и их широким применением на практике, прошло 20 лет.

Окно в 20 лет

Могла ли советская криптография как-то использовать это время и составить конкуренцию американцам в гражданской криптографии? Вспоминать об этом сейчас, столько лет спустя, довольно не приятно: вся советская криптографическая синусоида почти всегда шла следом за американцами, но никак не впереди их. Упущены были время, возможности и деньги.

Первые 10 из этих 20 лет – осторожное осознание преимуществ асимметричной криптографии и недоверчивость к ней. «Коварные американцы наверняка знают что-то, чего не знаем мы. Это криптографическая провокация!» Все идеи, вся теория – всё открытое, ничего секретного в алгоритмах асимметричной криптографии нет. Если часто трубить об американской провокации, то проблем можно получить несметное количество. Где доказательства, что это провокация? Что же советским

криптографам мешает узнать про то, что знают американцы? Почему американцы смогли создать асимметричную криптографию, а мы нет?

Советское криптографическое начальство быстро осознало, что особо широко пропагандировать американскую революцию в криптографии не стоит, несмотря на всю ее нетривиальность и практическую значимость. Так спокойнее, жили себе поживали только симметричной криптографией и горя не знали. Всё равно, реализовать асимметричную криптографию на советской элементной базе нельзя.

Во времена «дорогого Леонида Ильича» такая философия работала прекрасно. Но вот при Горбачеве появились в СССР эти проклятые персональные компьютеры западного производства и, естественно, встал вопрос: на советской элементной базе асимметричную криптографию реализовать нельзя, а как насчет западных персональных компьютеров? Там все можно.

Повторим: период, в течение которого советская криптография могла составить конкуренцию на мировых рынках, составлял 20 лет. Первые 10 – осознание идей, но вторые 10 – пора действовать! Создать на 10 лет раньше американцев «Комитет по изучению национальной политики в области криптографии», а в конце 80-х – начале 90-х гг. – свой OpenSSL, который можно было бы назвать, например, «Криптоцентр». То есть вторые 10 лет пора действовать, но кому и как?

Во время написания мной такого «Криптоцентра» внезапно начальник отдела собирает совещание.

– Товарищи, положение с продуктами в эту зиму будет очень трудным. В магазинах везде пустые полки. Мы договорились с подшефным совхозом послать им людей на помощь для уборки урожая. За это совхоз поставит нам картошку, капусту и морковь. Для их хранения будет задействовано бомбоубежище Спецуправления. Каждому отделению необходимо выделить по 2 человека для недели работы в совхозе.

Вместо «Криптоцентра» – жеребьевка, кто пойдет на неделю в совхоз. Вот так и использовалась полученная от американцев возможность выйти на мировой криптографический рынок. Не с программной реализацией асимметричной криптографии, а с капустой, картошкой, морковью и пустыми полками магазинов. Вторые 10 лет действовать не получилось.

Россия – не Америка!

В России довольно явный запрет на занятие криптографией появился примерно тогда же, когда в США опубликовали уже упоминавшийся выше

отчет «Роль криптографии в обеспечении безопасности информационного общества», в котором образованный для изучения этой роли Комитет призвал к «изменению национальной политики в области криптографии» с целью «дальнейшего экономического роста и лидерства ключевых отраслей промышленности США в глобальной экономике, в том числе (но не только) производстве компьютеров, программного обеспечения и средств связи».

В России никого ни к чему не призывали. В указе Президента РФ № 334 от 3 апреля 1995 года попросту запретили всякую деятельность, «связанную с разработкой, производством, реализацией и эксплуатацией шифровальных средств», без лицензии ФАПСИ.

Про «лидерство ключевых отраслей промышленности ... в глобальной экономике» в российском указе ничего не говорилось, не до этого было. В то (да и не только то) время одним из весьма эффективных российских бизнесов была выдача всяких разрешений, лицензий, сертификатов на все, что угодно. Главное – сесть на эту золотоносную жилу и делать на этом деньги без особого умственного напряжения. Вот тут-то уж хорошие офицеры намного важнее и полезнее, чем хорошие специалисты. А как сесть? С помощью своей административной власти и различной демагогии, направленной на запугивание рядового обывателя. Преступники воспользуются криптографией! Россия в опасности! Занятие криптографией без лицензии ФАПСИ равнозначно производству наркотиков! И еще миллион других разных причин, почему этот раздел математики требует контроля со стороны ФАПСИ.

Как же дошли до такого? Не сразу, примерно за три года. Тоже три года, как и американцы. Те изучали «Роль криптографии в обеспечении безопасности информационного общества» с 1993 по 1996. В России же с 1992 по 1995 готовили запрет на всякую криптографическую деятельность.

Так может быть лицензия ФАПСИ – это просто формальность? Нет, это деньги, нервы, поклоны чиновникам и масса всяких нормативных документов, которые можно понимать, как угодно.

Налицо закручивание криптографических гаек в середине 90-х. Если бы не резкая смена власти в стране, то, возможно, переход к гражданской криптографии тоже прошел бы плавно, по законам здравого смысла, примерно по американскому сценарию. Совершенно неожиданно к резкой смене власти в стране добавилась острая и немедленная потребность в гражданской криптографии. Я имею в виду проблему фальшивых банковских авизо, с которой столкнулся ЦБ РФ в 1992 году.

Кризис 1992 года

Об этой проблеме много писали в прессе того времени. Отмечу лишь некоторые основные моменты, играющие важную роль.

Кризис случился в 1992 году уже после преобразования 8 ГУ КГБ СССР в ФАПСИ и смены руководства.

В момент кризиса в ЦБ РФ не было достаточного количества персональных компьютеров, необходимых для организации с их помощью криптографической защиты во всех подразделениях ЦБ, поэтому требовалась разработка на отечественной элементной базе, выпускаемой на российском заводе.

Наиболее приемлемым вариантом отечественной элементной базы был портативный шифратор «Электроника МК-85 С».

Криптографическая защита нужна была очень срочно из-за огромного размаха мошенничества с помощью фальшивых авизо.

В середине 1992 года состоялось противостояние двух гигантов: ЦБ РФ и ФАПСИ. ЦБ РФ отвечал за финансовую безопасность России, а ФАПСИ – за безопасность связи. Указ Ельцина о монополии ФАПСИ на криптографию появился только в 1995 году, в 1992 году он только предполагался и обсуждался. Одной из обсуждаемых тем, к примеру, был запрет человеку, уволившемуся из ФАПСИ, в течение 5 лет работать по специальности, связанной с криптографией. В 1992 году монополия ФАПСИ на криптографию всего лишь предполагалась по умолчанию, но не была нигде закреплена законодательно.

Как делать деньги на криптографии? Что-то продавать. А что? Вариантов два: либо программную реализацию криптографических средств, либо аппаратную. Поскольку программная реализация обычно раз в 10 дешевле аппаратной, то основной задачей монополиста в области криптографии является продажа аппаратных средств.

Что было в середине 1992 года? Стандарт симметричного шифра ГОСТ 28147-89. Аппаратная реализация этого стандарта в виде отдельной платы «Криптон».

Мне представляется, что мог быть следующий диалог между ЦБ РФ и ФАПСИ.

– У нас проблема. Появилось множество фальшивых авизо. Хотелось бы получить для банковских авизо надежную криптографическую защиту от их подделки.

– В настоящий момент времени единственным средством криптографической защиты, отвечающим российским стандартам, является плата «Криптон».

В ЦБ РФ рассуждали так:

– плата «Криптон» может использоваться только

внутри персонального компьютера;

- персональных компьютеров в большинстве расчетно-кассовых центров (РКЦ) нет, и оснащение ими требует большого количества времени и денег;

- плата «Криптон» осуществляет шифрование информации, а для банковских альбомов нужна выработка кода подтверждения достоверности (КПД) в цифровом виде и небольшой длины, до 10 цифр.

Предположим, продолжили рассуждения в ЦБ РФ, что разработкой системы криптографической защиты банковских альбомов будут заниматься официальные структуры ФАПСИ. Как работают в этом случае типовые советские структуры, было известно очень хорошо: совещания, план, социалистическое соревнование, непредвиденные трудности, срыв сроков. Разработка и внедрение системы криптографической защиты банковских альбомов в этом случае растянулось бы года на 2 – 3. А мешки с деньгами, полученными по фальшивым альбомам, самолетами «Аэрофлота» возят в Чечню уже сейчас. Что делать?

Все подробности нетривиальной операции «Маркант», которая решила эту проблему, я изложил в одноименной книге [2], которая доступна на сервисе «Литрес». Сейчас попробую подвести некоторые ее итоги.

Операция «Маркант» была проведена ЦБ РФ с начала сентября по 1 декабря 1992 года. В результате ЦБ РФ получил надежную криптографическую защиту банковских альбомов, проработавшую в ЦБ свыше 30 лет. 1 декабря 1992 года, в день, когда в ЦБ РФ заработала эта система защиты, курс доллара по отношению к рублю упал на 8%.

Трудно было придумать более удобный момент для реального зарождения независимой от спецслужб гражданской криптографии в России! Американцы в 1993 году создали специальный комитет для «изучения национальной политики в области криптографии», 3 года ее изучали и в конце концов в 1996 году сделали вывод о том, что «преимущества широкого распространения криптографии перевешивают его недостатки». В России за 3 месяца, вопреки ФАПСИ, ЦБ РФ получил надежную криптографическую защиту банковских альбомов. На практике сразу стало всем ясно, что криптографическая защита – весьма эффективное средство в банковской сфере, что те преимущества, которые дает криптография, намного выше, чем издержки, связанные с ее первоначальным внедрением и сопровождением эксплуатации.

Операция «Указ»

В ФАПСИ решили: надо проводить свою операцию под названием «Указ», только уже более

тривиальную, административную, и разрешить заниматься криптографией только под контролем ФАПСИ. Эта идея витала в воздухе уже давно, но операция «Маркант» дала огромный импульс к ее практической реализации.

Так появился Указ Президента РФ № 334 от 3 апреля 1995 года. Он определял криптографическую политику в России, но, в отличии от США, «комитетов по изучению национальной политики в области криптографии» в России не создавали.

Прошло 30 лет – срок немалый для того, чтобы подвести и сравнить итоги обеих этих операций. Первая, которую проводил ЦБ РФ, позволила в чрезвычайно короткий срок создать и запустить в эксплуатацию надежную систему защиты банковских альбомов, которая проработала в ЦБ РФ свыше 30 лет. Вторая, которую проводило ФАПСИ, привела к монополизму, прямолинейности, безынициативности, отсутствию новых идей для криптографии в России и безусловному проигрышу России США в области гражданской криптографии в мире.

Сравнивая, по традиции, Россию и США, можно сделать такие выводы.

В США национальная политика в области криптографии направлена на то, чтобы сделать гражданскую криптографию максимально доступной и удобной для широких слоев населения, а также максимально независимой от спецслужб.

В России операция «Указ» определила национальную криптографическую политику в России, направленную на то, чтобы вся монополия на гражданскую криптографию принадлежала спецслужбам. Доступность и удобство гражданской криптографии в России для широких слоев населения отошли на второй план.

Монополизм

Коротко об электронной подписи (ЭП). Полагаю, что интересующийся читатель уже заглянул в Википедию и представляет, что это такое. А вот разработчики постановления Правительства РФ № 313 от 26 апреля 2012 года с названием из 64 слов заглянули в Википедию или куда-то еще забыли.

В этом постановлении говорится о криптографических средствах. Что такое эти самые средства? Ведь в постановлении говорится о лицензировании деятельности по разработке, производству и распространению криптографических средств. Ограничения очень строгие. Но что конкретно ограничивается? Все, что угодно. По крайней мере в этом нормативном документе в разделе 2 – что относится к шифровальным (криптографическим) средствам – название пункта в) подраздела 2 есть – «сред-

ства электронной подписи», а содержания – нет. Что считать электронной подписью и ее средством? Отсканированный файл с текстом документа и отсканированным отпечатком пальца под ним можно считать ЭП? Тогда сканер – средство электронной подписи.

Если в разработке «нормативной базы», которая появилась в России в качестве «нормативной базы» для гражданской криптографии, принимали участие математики – выпускники IV факультета Высшей школы, то мне стыдно за них. 5 лет лучшие преподаватели Советского Союза учили их математической точности определений, безупречной логике доказательств, ясности и четкости изложения, преданности криптографии, принципиальной гражданской позиции. И всё это – для составления подобной «нормативной базы»? Ну а для чего потребовалась криптографическим «верхам» России такая «нормативная база», вызывающая больше усмешек, чем уважения? Ответ на этот вопрос уже вполне логичен – для монополии на криптографию.

В 1995 году о выходе российской криптографии на мировые рынки и о конкуренции американцам речи уже не шло. Лет 10 назад, где-то в 1985 году об этом думать надо было. Кое-кто скажет, что и тогда это было нереально. В ответ я приведу пример одного выпускника IV факультета ВКШ, который успешно вышел на международные рынки и начал процесс такого выхода как раз где-то в середине 80-х. Он вышел с программным обеспечением, но не криптографическим. Его фамилию знают все.

В 1995 году в России «верхи» думали уже исключительно о российском рынке криптографических продуктов: как максимально оградить его от «не своих». Забор из всяких разрешений, лицензий и сертификатов – вещь хорошая. Но, как показывал пример с персональными компьютерами, если «не свои» создают что-то удобное, полезное и намного лучше российского, то есть большая вероятность, что и использовать в России начнут это удобное и полезное несмотря на то, что оно не российское. Значит, к забору нужно добавить что-то еще, чтобы у «не своих» пропала охота возиться с российским рынком и создавать криптографические продукты исключительно для него.

На мой взгляд, так появилась идея российских криптографических ГОСТов.

Российские криптографические ГОСТы

С 1985 года и по сей день во всем мире фактически применяются только три способа организации асимметричных криптографических преобразований. Про DH и RSA я уже упоминал, есть еще и пред-

ложенный в 1985 году американцами Нилом Коблицем и Виктором Миллером способ ECC – Elliptic Curve Cryptography. Прочитав про ECC в Википедии, легко понять, что с точки зрения практической реализации ECC похож на DH. Для зашифрования в нем используется пара «секретный ключ отправителя – открытый ключ получателя», в то время как в RSA – только открытый ключ получателя, за что RSA стал так популярен во всем мире. Все три способа – DH, RSA и ECC – чисто американские идеи, российского в них практически ничего нет.

Российские криптографические власти при выборе способа асимметричных преобразований для ГОСТ рассуждали, видимо, следующим образом.

Если сделать российским стандартом, связанным с асимметричной криптографией, RSA, то на наш рынок придут мировые гиганты типа Microsoft, которые запросто задавят всех наших отечественных монополистов уже тем, что RSA включен в криптопровайдер наиболее популярной операционной системы мира – MS Windows.

ГОСТ – российский, а способ ECC – американский. В RSA практически никаких параметров нет, а в ECC есть параметры, определяющие эллиптическую кривую. Своего, чисто российского способа, придумать не смогли, так придумаем свои параметры специально для России, с которыми Microsoft и ему подобные мировые гиганты связываться (включать в свою операционную систему) не захотят – рынок мал, весь под контролем спецслужб.

Замечу, что большинство простых обывателей о криптографии, как правило, имеют весьма туманное представление, а об RSA и ECC – тем более. Надо подкинуть обывателям какое-нибудь красивое словцо, чтобы вызвать у него страх и трепет перед российской асимметричной криптографией, основанной на американских алгоритмах ECC. Это уже задача не для хороших специалистов и даже не для хороших военных. Это задача для чиновников.

Криптографическая гигиена

Такое слово придумали: «квалифицированная». ЭП, оказывается, может быть не только подлинной или неверной, но еще и квалифицированной.

Крупные западные компании, например Adobe, в своих программных продуктах уже используют очень удобный и понятный интерфейс для проверки подписей файлов типа *.pdf. Для того, чтобы его использовать, нужно просто «соблюдать криптографическую гигиену», т.е. использовать общепринятые во всем мире криптографические алгоритмы.

Вот вам и «усиленная и квалифицированная»! Зачем и кому понадобилась эта игра слов? Лучше

бы соблюдали «криптографическую гигиену», т.е. подписывали полис ОСАГО с помощью RSA, а не «усиленной и квалифицированной».

Но ФАПСИ сделает мне справедливое замечание: ведь можно же адаптировать, например, Windows под российские ГОСТы! Тогда, возможно, и Adobe будет по-другому относиться к тем, кто не соблюдает «криптографическую гигиену».

Можно, не спорю. «Щедрая» душа – Microsoft – разрешил «кроить» свою самую популярную в мире операционную систему. Другие, например Apple, к своей IOS отнеслись более трепетно, и все жизненно важные файлы IOS подписали. Если подпись будет неверной, то и работать не будет. А MS разрешает использовать различным Самоделкинам свои так называемые kernel drivers, из-за которых весь Windows может кое-что у себя изменять. «Синий экран смерти» – знакомое понятие.

Аксиома программиста: программ без ошибок не бывает. Чем сложнее программа, тем больше вероятность допустить в ней какую-нибудь незаметную с первого взгляда ошибку, которая может привести к непредсказуемым последствиям. Что уж говорить про целую операционную систему Windows! Постоянно MS отлавливает в ней какие-то баги, выпускает к ним различные заплатки и обновления. Это – неизбежный процесс, если вспомнить философию, то Windows без багов – это как абсолютная истина в философии, вещь недостижимая. Могут быть достижимы только отдельные относительные истины (версии), но со временем появляется что-то новое, и процесс продолжается по спирали.

Действительно, Windows можно адаптировать под российские ГОСТы, добавив в него самодельный kernel driver. С оригинальным Windows ежедневно работают миллионы пользователей, он много раз проверен, но все равно, MS регулярно выпускает какие-нибудь обновления и модификации.

А сколько людей сейчас реально работают с Windows, адаптированным под российские криптографические ГОСТы? На несколько порядков меньше, чем с неадаптированным. И главный вопрос: зачем нужна адаптация замечательной OS Windows под российские криптографические ГОСТы? Какие у них преимущества перед мировыми криптографическими стандартами?

Россия против RSA

«Национальная политика в области криптографии должна преследовать следующую цель: широкая доступность криптографии для всех законных элементов ... общества» – из американского отчета «Роль криптографии в обеспечении безопасности

информационного общества». В нем сначала речь шла только об американском гражданском обществе. Но вскоре многие развитые страны тоже пришли к подобным выводам: нужна широкая доступность криптографии для простых законопослушных граждан. Ключевыми здесь были слова «широкая доступность криптографии».

И пошло развитие российской гражданской криптографии с середины 90-х годов по своему особому криптографическому пути, и неважно, удобен ли он пользователям.

Прошло почти 30 лет. Современное значение RSA в мировой общепринятой гражданской криптографии я бы сравнил с разъемом USB в бытовой электронике. Разъем USB принят всеми, удобен, широко распространен, легко доступен и т. д. К примеру, мышка к компьютеру практически всегда сейчас подключается через USB. А ведь были и другие времена, где-то до середины 90-х, когда мышку подключали через OS/2 (на компьютерном жаргоне – «полуось»). Но со временем USB полностью победил «полуось» в первую очередь потому, что этот разъем удобнее и более широко функционален. А что будет, если какая-то спецслужба в России сейчас запретит USB, велит всем перейти на «полуось» и создаст свой «свечной заводик» по производству «полуосей»? Трудно себе такое представить.

ФАПСИ и SSL

Вечно живую российскую аксиому про то, что строгость российских законов компенсируется необязательностью их исполнения, в российской национальной криптографической политике можно сформулировать так: RSA запрещен, но если очень надо, то можно.

А кто сказал, что RSA в России запрещен? Что сказано про это в нормативной базе криптографических запретов?

Читаем п 2. основного документа этой нормативной базы – Указа Президента РФ № 334 от 3 апреля 1995 года.

п. 2. Запретить использование государственными организациями и предприятиями в информационно-телекоммуникационных системах шифровальных средств, включая криптографические средства обеспечения подлинности информации (электронная подпись), и защищенных технических средств хранения, обработки и передачи информации, не имеющих сертификата Федерального агентства правительственный связи и информации при Президенте Российской Федерации...

Почти сразу же после выхода Указа № 334 ФАПСИ опубликовало разъяснения к нему в виде

статьи представителей ФАПСИ О.А.Беззубцева и А.Н.Ковалева «О лицензировании и сертификации в области защиты информации». И хотя с тех пор прошло уже почти 30 лет, данная статья вполне соответствует сегодняшним реалиям криптографической политики в России и Указа № 334 никто пока еще не отменял.

Теперь читаем в статье: «... цель заключается достичь до широкого круга специалистов суть и определение понятий в данной области так, как это отражено в нормативных актах ФАПСИ или официально трактуется Федеральным агентством (в случае наличия четких и ясных формулировок)» и « ... изделия, реализованные на базе собственных оригинальных алгоритмов, ранее не представлявшихся в ФАПСИ, а равно изделия, реализующие алгоритмы иностранной разработки, или импортные шифровальные средства на сертификацию не принимаются».

Сейчас, в эпоху проникшего во все сферы повседневной жизни Интернета, фактический запрет на «импортные шифровальные средства» воспринимается исключительно как неумение российских криптографических чиновников тогда, в 1995 году, предвидеть хотя бы на 5 лет вперед. Практически любой человек сейчас, пусть даже неявно, использует несертифицированные спецслужбами «импортные шифровальные средства». И не только человек, но и целые государственные структуры, по своей значимости никак не меньшие, чем прежде ФАПСИ. Например, портал «Госуслуги». Позволю себе и дальше криптографическую спецслужбу именовать, по привычке, ФАПСИ, хотя сейчас она сменила вывеску.

Для обращения к серверу «Госуслуг» используется протокол SSL, что означает, что весь трафик шифруется. Замочек, что появляется в левом верхнем углу экрана браузера пользователя, позволяет определить некоторые алгоритмы и их параметры, использующиеся при шифровании. Их описание приведено мной в книге «Маркант» [2], здесь дам только очевидный вывод: портал «Госуслуги» установил такой протокол обращения к своему серверу, при котором неоднократно используются несертифицированные импортные шифровальные средства.

Портал, на мой взгляд, получился замечательный в первую очередь потому, что его разработчики отличались здравомыслием.

Криптографическая защита информации, которая, безусловно, необходима при обращении к порталу Госуслуг, – это такое танго, которое надо танцевать вдвоем: криптографические средства, используемые для защиты, должны быть одинаковыми у сервера и у клиента. Клиентов в настоящее время

могут быть миллионы. Если поместить на сервер сертифицированные ФАПСИ криптографические средства, клиентов останутся единицы. Приказать миллионам клиентов использовать только сертифицированные ФАПСИ криптографические средства, конечно, можно, но какова вероятность того, что все эти миллионы клиентов подчинятся такому приказу?

Поэтому сервер Госуслуг выбрал SSL и несертифицированные ФАПСИ криптографические средства, поскольку у миллионов клиентов этого сервера на компьютерах или смартфонах эти криптографические средства уже есть, они включены в их операционные системы, стали мировыми стандартами, поскольку удобны, доступны, надежны и не зависят ни от каких спецслужб.

Глупости из прошлого

А делать надо было намного раньше.

Многие глупости, совершенные много лет назад, людивольно или невольно забывают. А потом разводят руками: ну надо же, как же так получилось? Но последствия этих глупостей могут неожиданно проявиться.

В начале 50-х советские «философы» объявили кибернетику «буржуазной лженеукой». Буржуазные кибернетики в Англии сделали «машину Тюринга» для взлома немецкой «Энигмы», что существенно влияло на ход боевых действий в годы войны. А что, кроме фактического запрета на развитие кибернетики, сделали советские «философы»? Их хотя бы пристыдили за то, что СССР в производстве компьютеров значительно отстал от западных стран?

В середине 70-х один знакомый мне генерал утверждал, что в первую очередь нужны хорошие офицеры, а потом уже хорошие специалисты. Это было время бурного научно-технического прогресса на западе. Уверен, что таких генералов-начальников в то время было немало. Западный научно-технический прогресс тех лет пролетел мимо СССР.

В середине 80-х в СССР наконец-то появились западные персональные компьютеры, которые никто не стал запрещать, как 10 лет спустя запретили «шифровальные средства импортного производства». С помощью ПК сравнительно легко создать «шифровальное средство». Считать ли его «импортного производства» или отечественного – никто таких вопросов не задавал. Появлялась реальная возможность выйти с криптографической продукцией на мировые рынки. Но для этого нужна была смена «национальной политики в криптографии», подобная той, которую провели американцы 10 лет спустя. Не решились, или просто не захотели.

Чтобы криптографическими стандартами пользовались миллионы пользователей, нужно создать свои компьютеры и операционные системы к ним, свои сертификационные центры по выдаче сертификатов для своей системы с открытым распределением ключей, сделать их максимально простыми и удобными для пользователей, чтобы они добровольно, без принуждения стремились ими пользоваться. Это и есть стартовая площадка для собственной национальной гражданской криптографии. Если невозможно её создать, нужно использовать то, что принято во всем мире – гражданскую криптографию по мировым стандартам.

У читателя наверняка возникнет вопрос: а тогда, в 1995 году при подготовке Указа № 334 и разъяснений к нему, разве нельзя было всё это предвидеть? По крайней мере, не было бы того, что видим сейчас на большинстве российских сайтов, использующих SSL: нельзя, но, делать нечего, приходится закрывать глаза. По большому счету, ничего страшного в том, что в SSL большинство российских сайтов используют несертифицированные ФАПСИ криптографические средства, нет. Это, как правило, очень надежные и удобные средства, используемые во всем мире. Но один эффект Указ № 334 все же дал. Это фактический запрет на криптографическую аутентификацию.

Криптографическая аутентификация

«Вам звонят из службы безопасности банка» – эту или подобную ей мошенническую фразу, я думаю, слышали многие из россиян. Цель, как правило, одна – заговорить доверчивого человека и выудить у него OTP – One Time Password – одноразовый код подтверждения банковского запроса, получив тем самым доступ к счету клиента.

OTP – это, как правило, 4 – 5 случайных цифр, которые банк посылает на мобильный телефон клиента и просит их ввести. Получив ответ от клиента, банк сравнивает отправленную и полученную комбинации и, если они совпадают, то аутентификация клиента считается выполненной успешно. Это, так называемая OTP аутентификация клиента, в ней нет никакой криптографии. Мошенники, от имени клиента, инициируют запрос к банковскому серверу на мошеннический перевод денег со счета клиента. Банковский сервер генерирует случайный OTP и посыпает его на телефон клиента. Мошенники звонят клиенту, заговаривают ему зубы и просят назвать полученный из банка OTP. Если доверчивый клиент верит мошенникам и сообщает полученный из банка OTP, то мошенники сообщают его банковскому серверу и он совершает мошеннический перевод денег со счета клиента.

Это очень краткое описание «OTP-мошенничества», т.е. мошенничества с помощью OTP, в реальной жизни мошенники придумывают много различных сценариев с той же самой целью.

OTP-мошенничество стало возможным и широко распространенным в России, потому что аутентификация с помощью только OTP – довольно слабая, все держится на 4 – 5 случайных цифрах, которые к тому же в явном виде сообщаются сервером клиенту. Что делать?

Вернемся в конец 1992 года, к проблеме «фальшивых азвизо». Быстрое и успешное решение этой проблемы стало возможным благодаря использованию криптографической защиты банковских азвизо. Мировой опыт, в частности, опыт Южной Кореи, показывает, что решить проблему «OTP-мошенничества» можно с помощью криптографической аутентификации (КА) клиента, когда сервер что-то посыпает клиенту и в ответ требует от него его персональную ЭП отосланного. Проверяется не совпадение посланного сообщения с полученным, а полученная ЭП отосланного сообщения.

Теперь давайте попробуем внимательно разобраться во всех плюсах и минусах КА перед OTP. Начнем с плюсов.

Основным орудием мошенников, как правило, является телефон клиента, на который приходит OTP. КА может вообще не использовать для аутентификации клиента его телефон, все общение клиента с банковским сервером может быть организовано с помощью только персонального компьютера или смартфона клиента. В России есть места, не покрытые сетями мобильной связи, а доступ в Интернет может быть обеспечен с помощью спутниковой связи. Российская мобильная связь может быть недоступна и в некоторых зарубежных странах, например, в Южной Корее.

Для КА необходим персональный секретный ключ (СК) клиента, который действует определенный и довольно долгий период времени. Клиент может вообще не иметь доступа к считыванию своего секретного ключа и даже не подозревать о его наличии. Например, СК может храниться в защищенном виде на специальной смарт-карте и осуществлять ЭП только внутри этой смарт-карты с помощью внутренней операционной системы смарт-карты без возможности считывания СК штатными методами внешними устройствами. Обмен с внешним устройством (например, компьютером) осуществляется по следующей схеме.

- компьютер подает на вход смарт-карты сообщение, которое нужно подписать;
- смарт-карта подписывает сообщение и выдает компьютеру только полученную подпись.

Компьютер или смартфон клиента посыпает запрос банковскому серверу на проведение транзакции, подписанный клиентом. Без ЭП клиента запрос на любые операции с его банковским счетом банковским сервером сразу же отвергаются. Если мошенник захочет послать запрос на банковский сервер о проведении мошеннической операции, то он должен его подписать с помощью ЭП клиента. Эта задача намного более трудная, чем заговорить клиента и выудить у него присланный из банка по телефону ОТР.

В итоге почти любой мошенник, столкнувшись с гарантированной криптографической стойкостью, неизбежно придет к выводу о безнадежности своего преступного занятия. Могут быть и исключения, упомянутые мной в книге «Невероятное», когда продвинутые мошенники пишут специальные программы для того, чтобы похитить СК из памяти компьютера клиента, но это трудно, дорого и небезопасно. В Южной Корее за несколько лет был всего один такой случай в самом начале использования в Интернет-банкинге КА, его потом лет пять вспоминала вся страна. Корейцы быстро и эффективно защитились от этой опасности, переведя секретный ключ клиента из оперативной памяти компьютера на защищенную смарт-карту.

Таким образом, при криптографической аутентификации получается гарантированно безопасная процедура опознания клиента. Так почему же в России онлайн-мошенничество стало так широко распространено? Почему фактически не применяется процедура криптографической аутентификации, а повсеместно используется намного более слабая процедура аутентификации с помощью ОТР?

Неадекватная современным информационным реалиям криптографическая политика в России – это только половина ответа. Неужели нельзя просто повсеместно ее игнорировать так же, как в случае с SSL?

К сожалению, для большинства законопослушных граждан, но к счастью для мошенников, нельзя. По крайней мере, практически никто не решается широко применять криптографическую аутентификацию. Помимо законопослушания, здесь есть много технических проблем. Для КА нужна сеть удостоверяющих центров, выдающих сертификаты открытых ключей по мировым стандартам. Опять же обращаясь к опыту Южной Кореи, они с этого начинали еще до широкого распространения КА. Я писал об этом во второй части книги [1].

Большая подготовительная работа – это основной минус КА перед ОТР. Но эта подготовительная работа выполняется только один раз перед запуском КА. После наладки и освоения КА ее исполь-

зование становится практически таким же простым, как и ОТР, что убедительно доказали корейцы на своем опыте.

Задачу постепенного перехода от ОТР к КА можно рассматривать как важную общегосударственную цель. Что необходимо сделать в первую очередь? Пересмотреть национальную политику в области криптографии, как это сделали американцы почти 30 лет назад, с учетом современных российских реалий.

Взвесить все плюсы и минусы запрета криптографических алгоритмов иностранного производства, дать разумные объяснения их обоснованности. На мой взгляд, большинство таких запретов приносят вреда намного больше, чем пользы. Широкое применение КА невозможно без использования RSA, теоретически запрещенного в России, но повсеместно применяемого российскими государственными сайтами в SSL.

Необходимо допустить возможность проведения криптографических исследований и выдачи результатов различных криптографических экспертиз экспертами, не зависящими от спецслужб. Результаты таких исследований и экспертиз должны давать право использования исследуемых и представленных на экспертизу криптографических решений, если не доказана их ошибочность.

Существующие в России национальные криптографические ГОСТы не отвечают потребностям современного информационного общества, они выбирались безо всяких конкурсов и сравнений с зарубежными аналогами. Постепенный и плавный переход в повседневной криптографической практике в России от ГОСТов к общепринятым мировым криптографическим стандартам мог бы стать одной из целей национальной политики в области криптографии в России.

Криптографическое просвещение широких слоев гражданского общества. Люди должны иметь доступ к знанию о том, что такое открытый и секретный ключи, чем они отличаются, правила обращения с ними, что такое симметричный и асимметричный алгоритмы шифрования, основные асимметричные алгоритмы: DH, RSA и ECC, чем они отличаются и где их применение наиболее обосновано, что такое хеширование и что такое ЭП, что нужно для осуществления ЭП (подписания), а что – только для ее проверки, что такое сертификат открытого ключа, кем и для чего он выдается, наиболее известные мировые и национальные центры по выдаче сертификатов и так далее.

В официальных документах следует избегать неоднозначных и малопонятных определений типа

«усиленная квалифицированная электронная подпись». Если вкратце резюмировать, то России необходима десакрализация криптографии, снятие с нее навязанной чиновниками таинственности и недоступности, объяснение криптографических методов и терминов наиболее простым и понятным языком.

Обязательно должны быть персонально названы люди, создавшие и поддерживающие в России с середины 90-х годов национальную криптографическую политику, противоречащую здравому смыслу. Вред, который принесла эта политика, сравним по своим последствиям с вредом от «философов», объявивших в начале 50-х кибернетику «буржуазной лжен наукой». Тогда это привело к длительной блокировке в СССР процессов, связанных с созданием и использованием вычислительной техники.

Российская политика в области криптографии фактически приостановила развитие и внедрение в повседневную жизнь новых криптографических методов и идей. По требованию спецслужб разрешено использовать только то, что объявлено российскими криптографическими ГОСТами. Лицензирование и сертификация криптографической продукции сопряжены с таким количеством бюрократических и часто просто необоснованных ограничений и запретов, что процесс получения сертификата от спецслужб на криптографическую продукцию чуть ли не превосходит по трудоемкости само создание такой продукции.

Требования спецслужб зачастую значительно снижают эксплуатационные характеристики, а выгода от выполнения этих требований бывает весьма неочевидна. Например, сравнение многих параметров операции хеширования, таких как скорость, объем требуемой памяти, сложность реализации и некоторых других. С помощью наиболее известных мировых стандартов хеширования практически всегда возможно создать более оптимальную реализацию операции хеширования, чем с помощью российского алгоритма «Стрибог», объявленного национальным российским стандартом и признанного спецслужбами единственным возможным в российских криптографических разработках.

Какие нужны изменения?

Предсказать процесс изменения национальной криптографической политики в России невозможно. Но дать некоторые субъективные рекомендации можно. Первое, что нужно сделать – это публично заявить о таком изменении в связи с тем, что применявшаяся с середины 90-х годов в России национальная политика в области криптографии оказалась неэффективной и контрпродуктивной.

Публично отменить все запреты в области криптографии, введенные с подачи спецслужб и в основном в их интересах. А затем создать российский «Комитет по изменению национальной политики в области криптографии». Отличие от подобного американского комитета – в одном слове: американцы создавали комитет по изучению национальной политики, а в России за 30 лет эту политику уже так хорошо изучили, что теперь однозначно требуется её изменение.

Желательно, чтобы этот комитет исследовал все случаи, где контроль спецслужб за криптографией действительно необходим, а где нет.

Если какой-нибудь фермер из Тверской области держит свои деньги в Сбербанке, то использование тех или иных криптографических методов во взаимодействии такого клиента с банком касается только клиента и банка, спецслужбы к этому не должны иметь отношения. Эти вопросы могут отражаться в договоре клиента и банка об обслуживании. Если клиент и банк согласятся использовать RSA и иные международные криптографические стандарты, спецслужба не должна в это вмешиваться и давать тем самым повод к принудительной монополизации рынка.

Если спецслужба, оправдывая монополизацию, начинает общие рассуждения про безопасность, защиту от преступников, мировой опыт, то пусть приведет конкретные аргументы, чем российские криптографические ГОСТы лучше общепризнанных мировых стандартов. И при этом не забывает о последствиях монополизации, из-за которой простые пользователи с криптографией вообще боятся связываться, что уже привело к тотальному мошенничеству, захватившему Россию вследствие слабой ОТР-аутентификации.

Целесообразно, чтобы Комитет рекомендовал меры государственной поддержки изменения национальной политики в области криптографии. К таким мерам в первую очередь можно было бы отнести создание российских удостоверяющих центров по выдаче сертификатов открытых ключей для алгоритмов общепризнанных мировых криптографических стандартов, в первую очередь RSA. Необходима поддержка продвижения таких удостоверяющих центров на мировой рынок и включения сертификатов этих центров в качестве корневых в общепризнанные мировые операционные системы, такие как Windows, Linux, iOS, Android и пр.

Желательно, чтобы комитет рекомендовал меры широкого криптографического просвещения всех простых граждан России, ввел основы криптографии в качестве обязательной дисциплины для изучения

в ВУЗах и колледжах информационной направленности, делая при этом упор на то, что криптография, как специфический раздел математики, не может подвергаться каким-либо ограничениям и лицензированию, что разработка и применение известных криптографических программ с любыми криптографическими алгоритмами являются обычными научно-исследовательскими задачами, которыми

может заниматься любой человек. Если какому-либо ведомству нужны ограничения на используемые этим ведомством криптографические методы и алгоритмы, то такие ограничения могут быть наложены только внутриведомственными нормативными документами.

СПИСОК ЛИТЕРАТУРЫ

1. Масленников М.Е. Криптография и свобода. URL: <https://www.litres.ru/book/mihail-evgenovich-maslennikov/kiptografiya-i-svoboda-71369458/chitat-onlayn/?ysclid=mhkjeuws3771614439>
2. Масленников М.Е. Маркант. URL: <https://www.litres.ru/book/mihail-evgenovich-maslennikov/markant-71339377/?ysclid=mhkitlv3ec117104854>

От редакции

В уходящем году было опубликовано интересное учебное пособие для начинающих специалистов по защите информации «Начала технической защиты информации» (Типография «Вишневый пирог», 2025. – 326 с.).

Автор – Светлана Валерьевна Конявская-Счастная – известный специалист в области защиты информации, а особенно – в области дидактики и методологии преподавания дисциплин этого цикла, позиционирует ее как учебное пособие для тех, кто планирует быть разработчиком средств защиты информации.

Аннотация книги дает представление не только о том, кого автор видит своей аудиторией, но и о том, как он (точнее, она) видит свою задачу, а именно – дать читателю почувствовать себя в роли и на месте бояршина.

Заметим, что это не самый распространенный посыл для учебной литературы. Как правило, цель заключается в том, чтобы дать необходимый объем знаний, который затем можно проверить на предмет качественного усвоения в стиле «на любой вопрос дам любой ответ». Однако книга «Начала технической защиты информации» не только сообщает полезные сведения (при известной настойчивости их можно разыскать самому и без учебника), сколько рассказывает о том, как мыслит безопасник в конкретных ситуациях. Читатель, в свою очередь, может попробовать определить, насколько ему близок этот образ мысли.

Нельзя сказать, что в этом учебном пособии отсутствуют признаки «нормального» учебника. В нем даются почти все необходимые определения, ключевые понятия системно соотносятся между собой и задаются связи, в тексте есть множество отсылок к различной литературе (в том числе и к дополнительной литературе по тому или иному вопросу), в конце каждой главы приведена библиография.

Особенно нашу редакцию порадовало, что есть целая глава, отведенная под «байки» – рассказы об истории создания различных решений, и глава, в которой подвергаются сомнению кажущиеся незыблыми устои (факторы аутентификации), а также отсутствие ответов к вопросам в конце учебника.

Всё это нацелено, по-видимому, на то, чтобы передать читателю прежде всего некоторый образ мыслей, близкий автору, а не на то, чтобы срочно подготовить его к экзамену по конкретной дисциплине. Какой именно это образ мыслей или посыл, вероятно, разные читатели ответили бы по-разному, однако нам кажется, что ключевой контрапункт – это приоритет здравого смысла и позитивного мышления над любыми устоявшимися и не всегда правильными привычками в сфере информационной безопасности. Книга призывает всегда задавать себе и другим вопрос «Почему?», и не успокаиваться, пока не будет найден ответ.

Насколько материал полезен и удобен для погружения в специальность – вопрос дискуссионный. К этому же дискуссионному аспекту необходимо отнести и авторский дискурс. Сама формулировка названия – «Начала технической защиты информации» (вместо более ожидаемых «основ» или «введения») – уже намекает на то, что под этим заглавием может оказаться текст, чтение которого потребует некоторых ментальных усилий.

Текст, действительно, не лишен личной окраски и авторской интонации, которая, вероятно, кому-то может не понравиться, но нам пришлась по вкусу. Примеры, приводимые автором, на первый взгляд могут удивить, однако едва ли стоит оставлять освоение по-настоящему сложных тем полностью на ответственность читателя. Поскольку не всегда можно проверить, что материал понят им и не будет бесследно забыт после того, как он закроет книгу.

При этом задача сделать материал доступным, возможно, и не решена окончательно и безусловно. Вот что рассказали нам об опыте использования этого учебного пособия в преподавании курса на кафедре защиты информации Физтех-школы радиотехники и компьютерных технологий МФТИ.

Студентам было задано на дом ознакомиться с главой про идентификацию и аутентификацию с тем, чтобы на паре обсудить прочитанное и прояснить все вопросы. Один из студентов заявил, что из этого текста совершенно ничего не понятно и он даже не может задать вопросы по материалу. Другой студент по собственной инициативе и без участия педагога разъяснил ему своими словами все положения главы, и спросил, что он, собственно, не понял из текста?

С одной стороны, ситуация показательная, поскольку для того, чтобы что-то объяснить другому, необходимо понять материал. А с другой стороны, студент, которому не подошел учебник, не изменил своего мнения и попросил задавать ему изучать проходимые темы по другим источникам.

Думается, что это очень красноречивый пример позиционирования книги «Начала технической защиты информации». Она вполне подойдет тем, кто любит при чтении находиться в диалоге с автором, но не тем, кого отвлекают примеры и раздражают аллюзии.



Начать с начала...

Неизбежный маршрут

Егор Федоров

Писатель, сценарист, драматург
Республика Беларусь



Меня зовут Платон, мне 23 года.

Так себе начало для рассказа, да? А если добавить, что во мне с рождения жили две сущности? А если добавить ещё, что вторую сущность мне подселили в качестве эксперимента тогда, когда я был эмбрионом? И что эту сущность сейчас бы назвали биологической нейросетью, биологическим искусственным интеллектом? Советская наука из-за незнания современных терминов называла её симбионтом.

Давайте приправим ещё всё это тем, что сущность эта была, кажется, женской.

Уже интереснее. Но при этом ничего не поменялось: меня зовут Платон и мне 23 года.

Представьте себе океан. Представили? Не смешите. Ни в одну человеческую голову не поместится океан. Это как представить себе меловой или кайнозойский периоды. Или миллиард лет истории Земли. Или миллиард долларов. Да, конечно, абстрактно можно представлять себе бюджет США в миллиардах. Но весьма умозрительно.

Так вот, представьте себе кусок океана километров хотя бы на десяток кубических. В этом куске океана плавает рыба дельфин. Хорошо, пусть не рыба, пусть просто дельфин. Он даже не плавает, а скорее находится в нем, в океане. Он подвешен в этом пространстве, как подвешивают кукол на длинных нитках, когда дают представление в кукольном театре. Только для дельфина ниток, конечно, никаких нет. Он покачивается долго, бесконечно долго.

Дельфин, который совсем не знает, что такое время, и есть я. Он покачивается до тех пор, пока не появляется второй дельфин, не похожий на него. Назовем второго дельфинихой. Та сильно поживее. Та больше увлечена окружающим миром. Она толкает, заигрывает с нашим первым дельфином и приглашает его в путешествие по остальному океану.

То, что я описал – моё первое воспоминание о моём существовании. И о существовании Симы – той самой дельфинихи, биологического ИИ, который внедрили в мой эмбрион. Потому она и Сима, от «симбионт».

С Симой я прожил почти 24 года своей жизни. А два месяца назад Сима ушла. Подробности я расскажу позже. Сейчас скажу, что с уходом Симы стала разрушаться и моя связь с прошлым. Ещё совсем недавно мое первое воспоминание о себе было ярким и наполненным множеством деталей, а сейчас остались лишь обрывки.

Когда я обратил внимание на то, что стал забывать о своем самом первом самоощущении, я стал отслеживать ситуацию и через какое-то время понял: а дело-то плохо. Теперь я теряю биты, байты и килобайты информации примерно так, как стираются сообщения в телеграмм-чате, если его настроить специальным образом.

Кто-то или что-то неумолимо стало вычищать, корректировать мою память. В этом ещё предстоит разобраться, а пока что я займусь сохранением, записью того, что осталось.

Итак, в утробе я был, пожалуй, рыбой. Это сейчас день сменяет ночь, а ночь сменяет день. Когда находишься в утробе, этого нет. В ней, как в океане, ощущаешь себя рыбой, у которой нет, не было и не будет в будущем никаких проблем, всё ровно, ей ни до чего нет дела, её всё устраивает. У меня не было даже и представления о том, что бывают какие-то проблемы или дела. Я просто был.

С появлением Симы моя жизнь поменялась. Я уже больше не висел куклой на нитках. Мы стали путешествовать по океану. И во время этих путешествий Сима стала меня учить. Для того, чтобы чему-то кого-то учить, прежде всего нужно выучить язык. Даже если этот язык – язык тела. Но если с языком тела мы справились очень быстро и я его усвоил на «отлично», то затем меня ждал более сложный язык – язык слов, понятий и образов.

Сначала Сима нарисовала в моей голове пустоту. Впрочем, нет, всё-таки не совсем пустоту, так как её очень сложно нарисовать. Даже в той, первой картине Симы в пустоте что-то было. Кажется, в ней мерцали звёзды. Память, как я говорил, уже подводит меня, и нужно поскорее записать всё то, что она ещё сохранила.

Пустота, в которой было мерцание звёзд, не пугала и не радовала, и мне показалось, что мы с Симой летим, потому что очень скоро в ней появилась наша Звезда, сначала маленькая и растущая, потом – огромная и необъятная. Затем возникла и вся солнечная система.

Мы облетели её, посмотрели на бури над Венерой, на огромный поток камней и льда колец Сатурна, на ледники Титана, на гейзеры Энцелады и вулканы Ио.

Времени как не существовало до Симы, так и не появилось с её появлением, поэтому я не знаю, как долго мы путешествовали. По ощущениям – очень долго. После чего Сима привела меня, наконец, к голубой планете.

– Это – наш дом, – сказала она и я почему-то её понял, хотя ни «это» ни «наш» ни «дом» в представлениях у меня не существовало. С другой стороны, понимает же как-то впервые младенец и «это», и «наш», и уж тем более – «дом»?

Мы висели в пустоте и смотрели, как вращается голубая наша планета и Сима спросила:

– Показать тебе, где мы сейчас?

– Конечно, покажи, – сказал я. Это было первое, что я сказал в жизни. Меня это очень удивило. Наверное, также, как удивляет маленького ребенка первый шаг: «Сам. Это сделал я сам». Сима взяла меня за руку и тут я понял, что у меня есть рука.

Голова шла кругом от новостей, но почему-то я сохранял ясность мысли и все эти новости воспринимал совершенно спокойно. Скорее всего из-за того, что рядом была Сима, чьё спокойствие распространялось и на меня. Это потом повторялось очень много раз в жизни. Сима взяла меня за руку, мы прошли облака и полетели к Земле.

Потом, ниже, были моря, леса, реки, поля, большие города и маленькие, были дороги, автомобили, поезда, аэропорты, вокзалы, башенные краны, башни из камня, в которых жили люди и башни из камня, в которых никто не жил. А потом мы снизились совсем уж к земле и пошли по ней пешком. Здесь я обнаружил, что у меня есть ноги. Тоже интересное обновление.

Мы пришли сначала к дачному поселку, потом к дачному участку, открыли калитку и вошли. Подошли к гамаку, на котором уютно лежала женщина с книгой. Около женщины стояла тарелка с абрикосами.

Сима сказала:

– Это тоже наш дом, – и показала женщину. – Мы с тобой здесь.

Она указала куда-то на живот женщины.

– Класс, – сказал я, и опять удивился тому, что сказал.

Итак, меня зовут Платон и мне по-прежнему 23 года. Все эти 23 года меня готовили в космонавты. Именно с этой целью мне был привит симбионт. Основная цель симбионтов – устранение потенциальных конфликтов на космическом корабле, который движется в космосе десятки, а может быть, и сотни лет.

Однажды в библиотеке космических исследований, в которую только люди причастные получали абонемент, я забрался в архив переписки теоретика космических перемещений профессора Антона Переверзева с Сергеем Королёвым.

«Звёздный Корабль медленно, команда за командой, со скоростью тактовой частоты Мироздания, то есть почти со скоростью света, переписывает себя в пространстве. Мы с Вами знаем, что он никогда не вылетит за пределы выделенной памяти. Но мы теперь знаем и как использовать Эйнсоф, Звёздные Врата. Но и после

этого нам много десятилетий ползти по новой системе до нужной планеты. И кто прилетит на неё? Несколько сумасшедших? Или Летучий Голландец с трупами убивших друг друга космонавтов?»

По-моему – исчерпывающе. Лишние конфликты на корабле совершенно ни к чему, и симбионты в первую очередь прививались для того, чтобы душить такие конфликты в зародыше.

Итак, симбионт был внедрен в меня для того, чтобы в команде космического корабля не случалось никаких ошибок человеческого плана, чтобы исключить ошибку человеческих отношений.

Знаю ли я механизм? Нет, но может быть, догадываюсь. Наверное, дело в том, что во время нашего взаимодействия влияние Симы на меня было настолько значимым, что она просто могла перехватить «управление» моим телом. Но как мне кажется, основные механизмы влияния были другими. Сима всегда запросто и без всякого насилия убеждала меня в нецелесообразности конфликта. Но возможно, что вмешательство было и на более тонких уровнях. Этого я не знаю, а знаю только, что были созданы сотни и тысячи таких симбионтов, как моя Сима, и все они были внедрены в эмбрионы.

Шел 1966 год. Это был, наверное, самый расцвет космической программы огромного и тогда мощного государства СССР, и планы по освоению космоса были грандиозными. Тем более, как я писал выше, профессор Переверзев вместе со своим НИИ открыл за год до этого Эйнсоф, или Звездные Врата – портал, через который можно было перемещаться, «прыгать» по Вселенной.

Тогда, в 1966 году, в космическую программу текли огромные средства, лучшие умы и самые фантастические технологии. Где же брали такие технологии? К примеру, откуда в то время могло взяться внедрение биологического ИИ – технологии, которая была утрачена спустя всего десятилетие. Также как и технологии производства собак с двумя головами или, допустим, технологии производства «вечных людей» – страшной технологии, благодаря которой «вечные» люди шли на суицид просто оттого, что больше не могли мучаться всё длящейся и длящейся жизнью. На это есть лемма Пошебякина, согласно которой со временем нейросеть в мозгу человека теряет исходную связность, и этот процесс завершается примерно через двести лет его существования.

Тогда же тысячам беременных женщин были внедрены тысячи симбионтов. Из них больше 50 процентов так никогда и не вышли на связь с фамильяром – владельцем тела, но другая половина всё-таки вышла. Нас, участников программы – сначала младенцев, а потом воспитанников детсадов и школ – не держали в специальном вольере. Мы ходили в обычные московские детские сады и школы. Но даже в этих обычных детских садах и школах я встречал таких же носителей Симы, которую носил в себе я.

Это были единицы, но мне кажется, что масштаб был большой. Кроме той функции, о которой я сказал, Сима обладала ещё массой полезных для меня качеств. К примеру, родовая память. Она умела к ней обращаться и предупреждать те недуги и проблемы, которые могли меня ждать в скором и даже более отдаленном будущем. В космосе не нужны психически неуравновешенные и больные люди, а нужны здоровые.

Были у Симы и другие функции – обучающая, развивающая, развлекающая – но они играли второстепенную роль.

Итак, меня зовут Платон и мне 23 года. Я родился в 1967-м году. То есть сейчас, как несложно сосчитать, 1990 год. И сейчас, в этом странном 1990-м году мы все, советские люди и учёные, которые готовились лететь в космос и двигаться в светлое будущее, получили вместо звезд икеевские фанерные табуретки. Как самый главный в жизни приз.

Когда всё катилось к этому – а со светлым будущим и космосом всё стало примерно понятно года эдак три назад – я часто стал задумываться: а что же случилось? Почему? Кристально ясно, что межзвездный проект потух вместе с тем, что у страны СССР, которая придумала этот грандиозный проект, начались грандиозные же проблемы.

Но мне не давал покоя вопрос: где здесь телега, а где – лошадь? Другими словами, страна СССР завершала своё существование из-за космического проекта или космический проект завершал своё существование из-за кончины СССР? И я пошел искать. Иду искать.

Мне кажется, здесь, на Земле, Сима не боялась никого и ничего.

Впрочем, это вполне закономерно: чего может бояться на Земле бестелесная и бессмертная сущность? Даже в те моменты, когда Сима «перехватывала» управление моим телом, рисковала она не собой, насколько я понимаю, а мной. И пару раз сильно рисковала.

Однажды, к примеру, меня стала терроризировать компания хулиганов, которая собиралась в детском саду, где моя мама работала заведующей. После школы я частенько заходил к ней за ключами. Конечно, мне могли

бы сделать мой личный комплект ключей, просто это был такой дополнительный ритуал, родительский контроль.

Мальчиком я в начальных классах рос несколько непредсказуемым, наверное, отчасти благодаря Симе. То она меня подначивала сходить на полдня вместо школы на аттракционы в парк, то мы с ней катались по Москве-реке на экскурсионном пароходе, то поставили себе цель обойти все в Москве кинотеатры и побывать там хоть на каком сеансе и даже достигли её. А порой мы просто с ней лазили по высоткам или по интересным домам. Например, я помню дом двухэтажных квартир, куда Сима меня привела на экскурсию.

Из-за всего этого родители меня слегка контролировали. Как поставленные на учет алкоголики ходят в милицию отмечаться, так я ходил к маме в детсад за ключами после своей первой смены в школе.

На выходе из сада, в беседке, меня стали задирать хулиганы. Чем-то я им не глянулся. А может просто промелькался: они там собирались почти каждый день, и я почти каждый день ходил за ключами. В общем, один раз что-то обидное стали говорить мне в спину. Второй. На третий они уже попытались стрясти у меня деньги – советскую мелочь. А в четвертый им понравилась моя кожаная куртка.

– Пригнись, – как-то весело и задорно сказала Сима. Один из хулиганов уже стоял передо мной и внимательно разглядывал мою куртку.

– Это как? – спросил я симбионта.

– Ну, попробуй сделать так, чтобы тебя здесь не было. Просто старайся ни о чем не думать, – ответила Сима.

– Как будто тебя нет.

Я постарался сделать, как она попросила и действительно на какое-то время из своего тела я просто исчез. То есть я был, но находился в каком-то мерцающем тумане, не в своем теле точно. Скорее всего, это было то пространство, где обычно обитала Сима. И выкинула меня в это пространство она.

Очнулся я уже в машине скорой помощи. Сима не уследила за всем происходящим. У одного из бандитов оказался нож, и пока она разбиралась с другими, он подобрался откуда-то сбоку и всадил его в меня. Тогда Сима умудрилась нейтрализовать и его. Как и другие, он уснул неприятным многодневным сном, видимо, с очень плохими сновидениями. Кто-то бы назвал это персональным адом. И только потом Сима моими ногами пошла к моей маме за скорой.

Ещё пару раз Сима «перехватывала управление» моим телом в высотках, чтобы походить по парапету крыши или чтобы попрыгать в классики с одной шатающейся плиты на другую.

Каждый раз после таких упражнений у меня, когда я возвращался, оставался след памяти. Скорее даже это был след впечатлений Симы и стресс моего организма от происходящего. Симбионт всегда просила прощения за то, что ей порой очень хотелось острых ощущений.

Так вот, на Земле она не боялась ничего и никого, но при этом я порой чувствовал, что она иногда испытывала страх. А может быть, даже ужас. Страх симбионта испытывала не перед тем, что видел я в своей реальности. А перед чем-то, что находилось в реальности, в которой была Сима. В один из таких приступов я спросил у Симы, что происходит. Она молчала несколько минут, потом сказала:

– Путешествуя в ментале, ты можешь встретить того, кому ты в лучшем случае безразличен.

– Но... зачем тогда там путешествовать? – спросил я.

– Пфф, – не уверен, что именно такой звук издала симбионт, может быть, что-то похожее. Но уверен в том, что это был звук презрения.

Она путешествовала в своём ментале, явно рисковала и страшилась этого риска. Чего она страшилась, она так и не сказала.

Когда я детально стал разбираться в том вопросе, который захотел решить, я понял, что то, чего она боялась, было как раз связано с тем, что забарахлила, закашлялась и зашаталась космическая программа в стране СССР. А потом и вовсе рассыпалась.

Тут было даже дело не в том, что я почувствовал это подсознательно. Я стал перебирать какие-то обрывки фраз и мыслей симбионта. Именно эти обрывки давали мне повод думать, что то, чего боялась Сима, было причиной того, отчего я не полетел в космос. Мы не полетели.

Но как подступиться к поиску ответа, чего боялась Сима, я не знал. Спросить уже было не у кого, сам я в это пространство погружаться не умел. Я вообще с трудом мог понять, что это такое. Да, я помню тот туман, в кото-



ром я себя чувствовал примерно ежиком из известного мультфильма. Я даже делал какие-то робкие попытки перейти туда самостоятельно. Не получалось. Да если бы и сумел, толку с этого было бы, кажется, немного.

Раз уж я рассказал вам про свою маму, расскажу и про отца. Он был ученым в звании полковника КГБ. К самой kontore глубокого бурения он отношения почти не имел. В первую очередь, он был, конечно, ученым. И, скорее всего, ученым, который имел отношение к проекту симбионтов. Потому что про мою Симу он знал.

— Чтобы лететь в Дальний Космос, что нужно, как ты думаешь? — спросил меня как-то отец после обеда. Я уже пил компот из чернослива. Выковыривал косточки. И не сильно задумался над ответом отцу.

— Я думаю, папа, что для этого нужен сам Дальний Космос... — ответил я как-то не особенно задумываясь. — А мы ведь закрыты в Солнечной системе. — Потом понял, что ответил, и прекратил пить свой компот. Я вспомнил Симу. — А еще, как мне... говорила...

— Не стесняйся, я знаю о ней.

— Сима говорит, что для этого еще нужны Цель, Воля и...

— И что еще?

— И она сама...

Всё-таки главное моё предназначение — полеты в космос — мне вколачивалось с самого детства. Следовательно, не искать причину, по которой ничего из этой затеи не вышло, я не мог.

Итак, я решил начать свои поиски с отца. Он был тем звеном, которое связывало меня с проектом симбионтов, с той самой библиотекой космических исследований и, возможно, с ответами на вопрос, почему всё пошло наперекосяк. Но в 1990 году, когда страна трещала по швам, как старая шинель, встретиться с ним оказалось не так просто. Он давно ушел в тень, в том числе и от меня, в какой-то закрытый институт на окраине Москвы, где, по слухам, все еще ковырялись в обломках былой славы.

Я раздобыл у мамы рабочий телефон отца и позвонил ему из телефонной будки на Тверской — мобильных телефонов тогда еще не было, а домашний мог прослушиваться.

— Папа, нужно поговорить, — сказал я.

— О чём? — не сильно приветливо уточнил отец.

— О Симе. О том, почему мы не полетели. О том, что вообще случилось с космической программой. И почему ушел мой симбионт.

Шизофреник обычно радуется, если Голоса у него в голове умолкают... Но я был совсем не рад, когда тем утром не услышал голос Симы. Я вообще долгие минуты без Симы не мог понять, что произошло. Так иногда вспоминаешь свои сны. И только минут пять спустя я вспомнил, как мы с ней прощались ночью, во сне.

Отец некоторое время молчал.

— Приезжай. Один.

— Хорошо, — ответил я и подумал, с какого это перепугу я приеду к отцу не один? А потом понял: он говорил о Симе.

Я теперь один.

Отец надиктовал мне адрес нашим старым шифром. Иногда люди особенным образом стучат в дверь, чтобы было слышно, что это свои. У нас же с папой с моих лет пятнадцати был свой особенный шифр.

Институт выглядел как заброшенный склад: серые бетонные стены, колючая проволока по периметру, и охранник с лицом, будто вырезанным из гранита. Такими изображают красноармейцев на барельефах.

Отец встретил меня в подвале, полном пыльных ящиков с папками.

— Привет, пап, — поздоровался я. Я особенно не рассчитывал, но он обнял меня.

— Привет, — сказал он, отстранившись. И пододвинул мне стул за столом, где работал. Перед ним стояло несколько стаканов с остатками чая.

Я присел. Папа сел напротив.

— Ну, добавить мне особенно нечего, — сказал я. — Она. Сима. Симбионт. Ушла два месяца назад.

Отец ничего не ответил.

— Пап, она ушла... — я запнулся. Потом вывалил свою догадку. — Она ушла потому, что мы никуда не полетели? Она поняла... Свою ненужность, что ли?

— Скорее, она поняла твою ненужность, — ответил отец.

Мы снова помолчали.

— И теперь ты теряешь память, да? — спросил меня отец.

– Да.

– Это нормально, – как-то туманно сказал он.

Папа закурил "Беломор" — старая привычка, которую мама ненавидела, — и начал рассказывать.

Оказывается, проект симбионтов родился не в лабораториях ЦНИИмаш и даже не в НИИ Переверзева.

Оказывается, эти технологии пришли извне.

– Ты читал «Сталкера»? «Пикник на обочине»? — спросил отец.

– Конечно, — я читал «Пикник на обочине» уже лет десять назад. Ещё даже до фильма Тарковского. — Только мне казалось, он был написан позже.

– Естественно позже, — недовольно ответил папа. — Он же не мог быть написан раньше событий.

И папа продолжил.

В 1965-м году открыли Эйнсоф — Звездные Врата. Это были не просто порталы в пространстве. Они открывали дверь в ментал. В то самое пространство, где обитала Сима.

– Ты наверняка понимаешь, что ментал — это не просто мерцающий туман, — сказал отец. — Это сеть. Бесконечная сеть разумов, сущностей, которые существуют вне времени и тела. Мы думали, что нашли короткий путь к звездам. Но на самом деле мы вторглись в их дом.

Отец немного помолчал, потом продолжил рассказывать.

СССР полетел бы в космос — планы были готовы. Флотилия кораблей, колонизация через Эйнсоф, симбионты для экипажей, чтобы избежать "Летучего Голландца" из письма Переверзева. Тысячи эмбрионов, таких, как я. Будущие космонавты, идеальные, без конфликтов. Но сущности в ментале... оказались нам не рады.

– Они — стражи, — продолжал отец. — Или, может быть, пастухи. Они видели, что мы не готовы. Наш эгоизм, наша жадность, наши войны — всё это могло распространиться по Метагалактике, как вирус.

Симбионты были их подарком, но и тестом. Они внедрили их, чтобы наблюдать и контролировать нас. Чтобы понять, можно ли нас пускать в этот другой мир. Знаешь, так приглядываются к чужому человеку — стоит или не стоит его вообще пускать на порог? А мы продолжали ломиться в Дальний Космос. И когда мы подошли слишком близко к запуску, в 1970-х, они вмешались.

Как? Не напрямую. Без взрывов ракет или саботажа — хотя и это было. Сущности работали тонко, через ментал. Они усилили "человеческий фактор": сомнения, бюрократию, коррупцию.

Потом был экономический кризис. Его усиливали страхами, которые сеяли симбионты. Да-да. Старшие имели влияние на симбионтов. Постоянное влияние, которым регулировали то, что происходит здесь.

– Старшие? — переспросил я.

– Сущности, которые находятся в ментале, — не очень довольно ответил отец. — Им было дано такое общее название.

– То есть это они сделали всё, чтобы космос для нас не состоялся? — спросил я. — Но зачем было нам давать этот шанс? Зачем было калечить столько судеб мечтой о Космосе, которая никогда не сбудется?

– Эх, — протянул отец. — Ты ещё Петrarку здесь начни цитировать. Или кого из других гуманистов.

Я понял, что действительно сказал глупость. Что я могу знать о морали этих Старших? Об их представлениях о мире? О нашем месте в этом мире?

Я немного подумал и спросил то, чего не понимал:

– Но почему... Почему Сима их боялась?

– Потому же, почему ты боишься Высших Сил, — ответил отец. — Только Сима своих создателей знала, а ты — нет.

Я сидел оглушенный. Загадка раскрывалась: СССР не полетел в космос не из-за денег или политики. Это была внешняя блокада. Сущности закрыли Эйнсоф, стерли технологии — вместе с этим пропали знания о "вечных людях" или двухголовых собаках. Они вытерли их из наших умов, как стирают чат в телеграмме. Страна рухнула, потому что без ментальной поддержки проект потерял импульс. Лошадь была менталом, а телегой был СССР.

Внезапно я решился.

– Папа. Ты не поможешь мне войти в ментал? — у меня была уверенность, что отец знает, как это сделать.



Отец покачал головой:

– Не стоит, – потом он заглянул в мои глаза и понял, что мне надо помочь. – Впрочем, если настаиваешь... Есть способ. Медитация, усиленная нашим устройством. Хотя, собственно, каким нашим. Тоже их подарок. Должны же были мы время от времени попадать в ментал. Оно открывает дверь.

Папа посмотрел на меня. Кажется, он заметил ту радость, что меня охватила.

– Но ты же понимаешь риски, – добавил отец. – Ты можешь запросто там застрять.

Я вспомнил ежика в тумане и мне сделалось отчего-то страшно, а смешно. Всю жизнь находится в этом мультфильме? Ну, и что теперь? Может быть, это будет интересно.

Я взял устройство — маленькую коробочку с электродами, похожую на старый энцефалограф. Дома я подключил его, закрыл глаза, вспомнил океан, дельфина, Симу.

И как будто бы нырнул.

Сейчас почему-то ментал не был туманом. Он был хаосом. Звезды, планеты, формы, меняющиеся как дым. Сначала я позвал Симу, надеясь неизвестно на что. Было понятно, что во всём этом огромном пространстве она меня не услышит. Но ничего лучшего я не придумал. Мало ли как тут, в ментале, распространяется мой этот зов?

Через некоторое время свет померк и вместо Симы пришла какая-то огромная, неясных очертаний, Сущность.

Старший.

– Почему вы остановили нас? – я собрал всю свою смелость, чтобы задать этот вопрос.

Голос, нет — мысль — заполнила всё: "Вы — дети. Играете с огнем. Ваши конфликты уничтожат не только вас. Мы дали симбионтов, чтобы научить эмпатии, единству. Но вы стали мечтать об экспансии и завоеваниях. Космос — не для войн. Когда вы вырастете, дверь откроется снова."

Тут я понял, что какая-то сила выдавливает меня из ментала. Наверное, так себя чувствует пробка шампанского в бутылке, успел подумать я и проснулся. Я посмотрел на будильник: в ментале я пробыл около трех минут. Хотя мне показалось, что прошло несколько часов.

Я лежал в тишине и думал о том, что СССР не полетел, потому что человечество не прошло тест. Нас заблокировали, вызвав цепь логичных и последовательных событий — от смерти Королёва до появления Горбачева и перестройки.

С каким-то ужасом и интересом я начал припоминать всю эту цепь и теперь понял, куда пропали шесть часов. Сущность показывала или рассказывала мне всю эту цепь. Зачем? Наверное, это произошло потому, что у меня был огромный, страшной силы импульс запроса внутри на эту тему. Это вечное «мечтайте осторожнее, мечты сбываются».

Сначала я мечтал о Космосе. А потом с не меньшей силой я стал мечтать о том, чтобы понять, почему не получилось. Но теперь я увидел где-то вдали луч надежды. Потому что из всех разговоров со Старшим я знал: маршрут неизбежен, и нужно только подождать.

Сима, если ты меня слышишь — спасибо. За океан. За звезды. За то, что ты была в моей жизни.

Итак, меня зовут Платон, мне 23 года.

И я всё еще жду полета.

Но теперь я знаю, куда лететь.

Рисунки заимствованы из общедоступного источника Интернет, не содержащего ссылок на авторство.

Приглашаем авторов к участию в журнале «Вестник современных цифровых технологий»

ИНФОРМАЦИЯ ДЛЯ АВТОРОВ

Редакция принимает материалы статей, соответствующие тематике журнала, содержащие новые научные результаты, не опубликованные ранее и не предназначенные к публикации в других печатных или электронных изданиях. Проводится независимое внутреннее рецензирование. Статьи в журнале публикуются бесплатно (объем – до 15 тыс. знаков), после получения одобрения Редакционного совета.

Для опубликования статьи в редакцию журнала необходимо направить по адресу a.shcherbakov@c3da.org, a.guazanova@c3da.org следующие материалы в электронном виде:

- рукопись статьи в DOC- и PDF-форматах;
- иллюстрации, предоставленные также и отдельными файлами в формате JPG или PNG с разрешением 300 dpi;
- сведения об авторах, содержащие фамилию, имя, отчество, ученые степень и звание, должность, место работы, контактные телефоны или E-mail;
- англоязычную информацию, содержащую название статьи, ФИО авторов, аннотацию и ключевые слова;
- редакция может запросить экспертное заключение о возможности публикации статьи в открытой печати.

ПОСЛЕДОВАТЕЛЬНОСТЬ МАТЕРИАЛОВ ДЛЯ ПУБЛИКАЦИИ:

1. шифр УДК (см. Справочник УДК) в левом верхнем углу;
2. название статьи (полужирным шрифтом по центру) не более 12 слов;
3. инициалы и фамилия автора (полужирным шрифтом по центру), к каждому автору - сноска, содержащая ученое звание, должность, название организации (без сокращений), e-mail;
4. Аннотация, излагающая суть работы и полученные результаты (5-7 строк);
5. ключевые слова (8-10 слов);
6. англоязычная информация по статье (по пп.2-5)
7. текст статьи с учетом указанных далее требований к его оформлению;
8. список литературы, оформленный по ГОСТ Р 7.0.5-2008.

Статья должна быть структурирована, т.е. должна включать разделы с названиями, кратко и точно отражающими их содержание, в том числе:

- введение, содержащее обоснование актуальности и краткий обзор проблематики;
- четкую постановку задачи исследования;
- описание метода решения задачи исследования;
- прикладную интерпретацию и иллюстрацию полученных результатов исследования;
- заключение, включающее обобщение и указание области применения полученных результатов, не повторяющее аннотацию и не ограничивающееся простым перечислением того, что сделано в работе.

С детальными требованиями к рисункам, таблицам, формулам, списку литературы, а также с примерами оформления статьи можно ознакомиться на странице Вестника <http://c3da.org/journal.html>.

Приглашается к сотрудничеству редактор для работы в редакции журнала по совместительству. Просьба направлять резюме по электронному адресу accda@c3da.org, info@c3da.org

ТРЕБОВАНИЯ К РЕДАКТОРУ:

- отличное знание русского языка;
- свободное владение ПК, в том числе специальными текстовыми и графическими программами;
- опыт работы в издательстве.

Высшее техническое образование и знание английского языка являются существенными преимуществами.

ОБЯЗАННОСТИ

Редактор:

- редактирует рукописи, принятые к изданию;
- оказывает авторам необходимую помощь по улучшению структуры рукописей, выбору терминов, оформлению иллюстраций;
- проверяет, насколько учтены авторами замечания по доработке, предъявленные к рукописям;
- подписывает отредактированные рукописи в печать.